DOCUMENT RESUME

ED 477 358 TM 034 920

AUTHOR	Rudner, Lawrence M., Ed.; Schafer, William D., Ed.
TITLE	Practical Assessment, Research and Evaluation, 2000.
INSTITUTION	ERIC Clearinghouse on Assessment and Evaluation, College Park, MD.; Maryland Univ., College Park. Dept. of Measurement, Statistics & Evaluation.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
ISSN	ISSN-1531-7714
PUB DATE	2000-12-00
NOTE	59p.; "Practical Assessment, Research & Evaluation" is an electronic-only journal covered on an article-by-article basis in "Current Index to Journals in Education" (CIJE). For the first 22 articles in volume 7, see ED 458 254. For articles 23 through 26, see TM 525 154-157.
AVAILABLE FROM	For full text: http://ericae.net/pare.
PUB TYPE	Collected Works - Serials (022) -- ERIC Publications (071)
JOURNAL CIT	Practical Assessment, Research and Evaluation; v6 n1-14 2000
EDRS PRICE	EDRS Price MF01/PC03 Plus Postage.
DESCRIPTORS	*Educational Research; Elementary Secondary Education; *Evaluation; *Item Banks; *Literacy; *Performance Based Assessment; Standardized Tests; Test Coaching

ABSTRACT

This document consists of articles 1 through 14 of volume 6 of "Practical Assessment, Research & Evaluation": (1) "Seven Myths about Literacy in the United States" (Jeff McQuillan); (2) "Implementing Performance Assessment in the Classroom" (Amy Brualdi); (3) "Some Evaluation Questions" (William Shadish); (4) "Item Banking" (Lawrence Rudner); (5) "Teacher Comments on Report Cards" (Amy Brualdi); (6) "Classroom Questions" (Amy Brualdi); (7) "Communicating Educational Research Data to General, Nonresearcher Audiences" (Gail S. MacColl and Kathleen D. White); (8) "Helping Children Master the Tricks and Avoid the Traps of Standardized Tests" (Lucy Calkins, Kate Montgomery, and Donna Santman); (9) "Strategies for Improving the Process of Educational Assessment" (M. Kevin Matter); (10) "Performance Assessment Links in Science" (Edys Quellmalz, Patricia Schank, Thomas Hinojosa, and Christine Padilla); (11) "The Nature of Evaluation Part I: Relation to Psychology" (Michael Scriven); (12) "The Nature of Evaluation Part II: Training" (Michael Scriven); (13) "How To Write a Scholarly Report" (Lawrence M. Rudner and William D. Schafer); and (14) "A Process for Evaluating Student Records Management Software" (Lisa Vecchioli). (SLD)

Volume 6, 2000 ISSN 1531-7714

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal

(Articles 1-14)

Practical Assessment, Research and Evaluation
is listed among the ejournals in education at the website for the
AERA SIG "Communications Among Researchers"

Practical Assessment, Research and Evaluation (PARE) is an on-line journal published by the **ERIC Clearinghouse on Assessment and Evaluation (ERIC/AE) and the Department of Measurement, Statistics, and Evaluation at the University of Maryland, College Park.** Its purpose is to provide education professionals access to refereed articles that can have a positive impact on assessment, research, evaluation, and teaching practice, especially at the local education agency (LEA) level.

Manuscripts published in *Practical Assessment, Research and Evaluation* are scholarly syntheses of research and ideas about issues and practices in education. They are designed to help members of the community keep up-to-date with effective methods, trends and research developments. While they are most often prepared for practitioners, such as teachers, administrators, and assessment personnel who work in schools and school systems, *PARE* articles can target other audiences, including researchers, policy makers, parents, and students.

Manuscripts to be considered for *Practical Assessment, Research and Evaluation* should be short, 2000-8000 words or about eight pages in length, exclusive of tables and references. They should conform to the stylistic conventions of the American Psychological Association (APA). See the Policies section of this web site for technical specifications and a list of suggested topics. Manuscripts should be submitted electronically to pare2@ericae.net. Articles appearing in *Practical Assessment, Research and Evaluation* also become available in the ERIC database through the ERIC Digest Series. Many articles published in *PARE* were previously published as part of the *ERIC/AE Digest Series.*

Practical Assessment, Research and Evaluation is listed among the ejournals in education by the Scholarly Publishing and Academic Resources Coalition(SPARC), and the website for the AERA SIG "Communications Among Researchers".

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal

Articles Pages

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

McQuillan, Jeff (1998). Seven myths about literacy in the united states. *Practical Assessment, Research & Evaluation*, 6(1). Available online: http://ericae.net/pare/getvn.asp?v=6&n=1. This paper has been viewed 15118 times since 11/13/99.

Seven Myths about Literacy in the United States

Jeff McQuillan,
California State University, Fullerton

Adapted from *The Literacy Crisis, False Claims, Real Solutions* (1998) by Jeff McQuillan, Portsmouth, New Hampshire: Heinemann.

Serious problems exist with reading achievement in many United States schools. However, much of the commonly accepted wisdom about the academic performance of United States students is false. The best evidence we have on the reading crisis indicates that no crisis exists on average in United States reading. The purpose of this article is to investigate seven of the most prevalent -- and damaging-- myths about literacy achievement in the United States.

Myth 1: Reading Achievement in the United States Has Declined in the Past Twenty-five Years

The best evidence on reading achievement in the United States comes from a national system of examinations established back in the late 1960s by the federal government to determine how United States schoolchildren were performing in a variety of school subjects. These exams, known as the National Assessment of Educational Progress (NAEP) are important barometers of educational achievement. They are given nationally to a representative sample of United States children.

When the test was first administered in 1971, the average reading proficiency score for nine year-old children was 208, for thirteen year-old children was 255, and for seventeen year-old children was 285. The results of the most recent administration of the test (1996) revealed that the average reading proficiency score for nine year-old children was 212, for thirteen year-old children was 259, and for seventeen year-old children was 287. These scores indicate that, despite a few minor shifts, reading achievement has either stayed even or increased over the past thirty years.

Myth 2: Forty Percent of U.S. Children Can't Read at a Basic Level

During the early years of the NAEP tests, the Department released only the raw scores for each age level on its 0 to 500 scale, with no designations of which score was thought to constitute "basic knowledge" or "proficiency." The designers of the NAEP test later decided that simply reporting the raw scores was no longer adequate in order to judge the progress of United States schools. The Department decided it would determine how well students were reading by establishing the minimum score constituting "below basic," "basic," "proficient," and "advanced" reading. The "basic" level for fourth-grade reading, for example, was fixed at a score of 208. In 1994, 40% of United States children scored below the "basic" cutoff of 208.

The problem with this approach lies in "objectively" determining where these cutoff points should be. Glass (1978), after reviewing the various methods proposed for creating "minimal" criterion scores of performance, concluded that all such efforts are necessarily arbitrary. Of course, such arbitrary cutoff points already exist in education and many other fields, but at least they are recognized as arbitrary and not given the status of absolute or objective levels of competence. In 1991, the General Accounting Office (GAO) examined the how the NAEP defined their levels of proficiencies and found their methods to be questionable (Chelimsky, 1993).

Myth 3: Twenty Percent of Our Children Are Dyslexic

Closely related to the previous misconception that 40% of our students read below the "basic" level is another portentous-sounding figure that indicates 20% of United States schoolchildren suffer from a "neuro-behavioral disorder" known as "dyslexia" (Shaywitz et al., 1996). The research most often cited to support this claim is drawn from the results of the Connecticut Longitudinal Study (CLS), a large-scale project funded in part by the National Institute of Child Health and Human Development (e.g. Shaywitz, Escobar, Shaywitz, Fletcher & Makuch, 1992; Shaywitz, Fletcher & Shaywitz, 1994). The CLS tracked over 400 students from kindergarten through young adulthood, periodically measuring their Intelligence Quotient (IQ), reading achievement, and mathematical abilities, among other attributes. CLS researchers measured "reading disability" by two methods. The first is what is known as "discrepancy scores," which represent the difference between a child's actual reading achievement and what would be predicted based upon his IQ. The idea is that if you have a high IQ but are poor at reading, then something must be wrong with you. The actual size of the discrepancy used in the CLS studies was that recommended by the United States Department of Education, 1.5 standard deviations. This 1.5 standard deviation figure was thus their "cutoff" score used to determine who was reading "disabled" and who was not. In any given year, a little less than 8 percent fall into the category of reading disabled, using the 1.5 cutoff.

Two important things need to be noticed about these results. First, and most importantly, the 1.5 standard deviation cutoff point is arbitrary. We could just as easily have used 1.25 or 1.75 or .5, each producing a different percentage of "neuro-behaviorally" afflicted children. Second, even the 8% have not been shown in this research to be "dyslexic," if by "dyslexic" we mean a "neurologically based disorder in which there is unexpected failure to read," the definition used by the CLS team (S. Shaywitz et al., 1992, p. 145; emphasis added). This is because, quite simply, no neurological measures were administered to these particular children. All that can be said from these findings is that around 8 percent of children in any given year will have a discrepancy of 1.5 standard deviations between their IQ and reading achievement, at least if they live in Connecticut.

Myth 4: Children from the Baby-Boomer Generation Read Better than Students Today

Some argue today's reading levels are dismal compared to those of the 1940s or 1950s. This evidence comes from a study of adult literacy levels, the National Adult Literacy Survey (NALS), which was given to a representative sample of United States adults in 1992 (Kirsch, Jungblut, Jenkins & Kolstad, 1993). McGuinness (1997) notes that those who learned to read in the mid-1950s to mid-1960s have higher reading scores than those of later generations.

Can we really measure the effectiveness of schools 40 years ago by how well their graduates read today? What about the intervening 30 years of reading experience and education? We should hardly expect the reading

proficiency of these adults to remain stagnant over time. Surely the reading scores of this group of 35-44-year-olds from when they were still enrolled in school are better indicators of how well they performed as children, since fewer intervening variables then exist to confound the results. We do, in fact, have reading achievement scores from a representative sample of children of this age cohort in the form of the high school NAEP scores from 1971 (for those who entered first grade in 1959 and were 38 at time of the NALS administration). Their scores are not much different than more recent graduates.

Myth 5: Students in the United States Are Among the Worst Readers in the World

What will come as most surprising to many people is how the United States compares internationally in reading achievement: Our nine-year-olds ranked second in the world in the most recent round of testing conducted by the International Association for the Evaluation of Educational Achievement (IEA); our fourteen-year-olds ranked a very respectable ninth out of 31. A dissenting opinion on just how well United States school-children perform over time and internationally is held by Walberg (1996), who argues that reading achievement has in fact declined since the early 1970s. Walberg compared the IEA scores from 1990-91 to the first IEA test given to 15 nations in 1970, with the scores from the two tests equated (Lietz, 1995, cited in Walberg). Walberg (1996) concluded that the scores did indeed decline, from 602 in 1970 to 541 in 1991 (using his adjusted scores).

Two problems exist with this analysis, however. First, it is not clear why the two IEA tests given 22 years apart should be preferred in measuring trends in United States reading performance over the United States Department of Education's own NAEP exam, which has not only been given more frequently (9 times since 1970), but was designed to be much more sensitive to a broader range of reading achievement (Binkley & Williams, 1996) than the IEA tests. Second, the IEA test has changed considerably since its first administration in 1970 (Elley, 1994). Unfortunately, the reanalysis of the scores upon which Walberg bases his comparisons is unpublished, making it difficult to know precisely how these "equated" scores were derived from what were markedly different tests.

Myth 6: The Number of Good Readers Has Been Declining

It has been claimed by some critics that the number of students "at the top" has been declining (e.g., Murray & Herrnstein,1992; Coulson, 1996). While it is true that the number of students scoring above 700 on the SAT did decline, the numbers were never high (2.3 percent in 1966, 1.2 percent in 1995). Also, the large demographic changes in United States schools over the past three decades have almost certainly had an influence on the scores. Bracey (1997) points out that the drops occurred primarily between 1966 and 1972, since which time the percentage of students scoring above 700 has remained stable. Moreover, two studies that have attempted to control for the significant demographic shifts in the test pool since the early 1950s have found that the average declines during the 1960s and 1970s were rather small (Bracey, 1997).

However, the most important point to keep in mind when discussing the SAT is that it is not a representative sample of United States high school students. It is a voluntary test that a large proportion of students takes in some states (e.g., New York) and hardly any students take in other states (e.g., Iowa). The NAEP tests, by contrast, are representative. They indicate no decline in the percentage of students who score at the highest levels. Little change has occurred in the percentage of high-scoring students at any grade level, with the percentage of thirteen-year-olds scoring at the top levels showing an increase over the past three decades.

Myth 7: California's Test Scores Declined Dramatically Due to Whole Language Instruction

In addition to finding a crisis where none exists, it has also become necessary to produce a guilty party to blame for our greatly exaggerated woes (Levine, 1996; Stewart, 1996): "whole language." The focus of these attacks has centered primarily on California, a state that at least nominally adopted a more "holistic" view of teaching language arts back in 1987. This supposedly led to a steep decline in reading scores.

Two points are at issue in the case of California and its reading crisis. First, did California's reading test scores really "plummet" (Stewart, 1996, p. 23) to record lows after 1987? Second, was this sharp decline attributable to the adoption of a reading curriculum in the state in 1987 (CRTFR, 1995), that emphasized reading books and decreasing (but not eliminating) phonics and skills instruction? It turns out that the answer to both of those questions is "no." The popular wisdom about California's decline stemmed mostly from the release of two sets of test scores: the 1992 and 1994 NAEP scores, and results of the state's own California Learning Assessment System (CLAS). In both the 1992 and 1994 state NAEP rankings, California fared rather poorly: In 1992, the state was in the bottom third, and in 1994, in the bottom quarter (Campbell, Donahue et al., 1996). Although Californian students clearly performed poorly compared to the rest of the nation, one must look at scores from both the beginning and the end of the time period in question to show a decline. Unfortunately, state-level NAEP scores are unavailable before 1992, and the tests are not equivalent to any other standardized reading measure. As such, the NAEP data cannot tell us anything about whether scores went up or down after the implementation of the literature-based curriculum. The only test score data available both before and after the implementation of the "holistic" 1987 Language Arts Framework are the California Achievement Program scores. However, there is no indication of dramatic drops or increases.

The second part of the argument used to promote a renewed emphasis on skills instruction was that whole language was the cause of California's (nonexistent) decline and (very real) low national ranking. Is a literature-based curriculum or whole language to blame? Another look at the 1992 NAEP data reveals that the answer appears to be "no." As part of the assessment, fourth-grade teachers were asked to indicate their methodological approach to reading as "whole language," "literature based," and/or "phonics." The average scores for each type of approach were then compared, and those children in classrooms with heavy emphasis on phonics clearly did the worst. Children in whole language-emphasis classrooms (reported by 40 percent of the teachers) had an average score of 220, those in literature-based classrooms had a score of 221 (reported by 49 percent of the teachers), and students in phonics classrooms (reported by 11 percent of the teachers) had an average score of 208 (NCES, 1994, p. 284).

Conclusion

Many things are wrong with United States schools. However, false crises and distorted views of student achievement can only distract us from the real concerns of parents, teachers, and policymakers. Instead, we need to have some understanding of what reading is and know some of the most important factors influencing reading achievement.

References

Binkley, M. & Williams, T. (1996). *Reading literacy in the United States: Findings from the IEA reading literacy study.* Washington, D.C.: National Center for Educational Statistics.

Bracey, G. (1997). *Setting the record straight: Responses to misconceptions about public education in the United States.* Alexandria, VA: Association for Supervision and Curriculum Development.

Campbell, J.; Donahue, P. et al.(1996). *NAEP 1994 reading report card for the nation and the states.* Washington, D.C.: US Department of Education.

Chelimsky, E. (1993). National Assessment Governing Board (NAGB)

Achievement Levels. Interim Letter Report. Washington, DC: General Accounting Office. (ERIC Document Reproduction Service No ED342821).

Coulson, A. (1996). Schooling and literacy over time: The rising cost of stagnation and decline. *Research in the Teaching of English*, 30, 311-327.

Elley,W. (1994). Preface. In W. Elley (Ed.), *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems* (pp. xxi-xxii). Oxford, England: Pergamon.

Glass, G. (1978). Standards and criteria. *Journal of Educational Measurement, 15*, 237-261.

Kirsch, I., Jungblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey.* Washington, D.C.: National Center of Educational Statistics.

Levine, A. (1996). America's reading crisis: Why the whole language approach to teaching reading has failed millions of children. *Parents, 16,* 63-65, 68.

McGuinness, D. (1997). *Why our children can't read and what we can do about it: A scientific revolution in reading.* New York: The Free Press.

Murray, C. & Herrnstein, R.(1992). What's really behind the SAT-score decline? *The Public Interest, 106*, 32-56.

National Center for Education Statistics (NCES) (1994). *Data compendium for the NAEP 1992 Reading Assessment of the nation and the states.* Washington, D.C.: U.S. Department of Education.

Shaywitz, S.; Escobar, M.; Shaywitz, B.; Fletcher, J.; & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *The New England Journal of Medicine, 326*(3), 145-150.

Shaywitz, S.; Fletcher, J.; & Shaywitz, B. (1994). Issues in the definition of and classification of attention deficit disorder. *Topics in Language Disorders, 14* (4), 1-25.

Shaywitz, B. et al. (1996). *The Yale Center for the Study of Learning and Attention: Longitudinal and neurobiological studies.* Dallas, TX: Paper presented at the Annual Meeting of IDA.

Stewart, J. (1996). The blackboard bungle: California's failed reading experiment. *LA Weekly, 18*(14), 22-29.

Walberg,H. (1996). U.S. schools teach reading least productively.

Research in the Teaching of English, 30, 328-343.

Descriptors: *Academic Achievement; College Entrance Examinations; Comparative Analysis; Dyslexia; Educational Research; *Educational Trends; Elementary Secondary Education; International Studies; *Literacy; *Mythology; National Surveys; *Reading Achievement; Trend An

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

Brualdi, Amy (1998). Implementing performance assessment in the classroom. *Practical Assessment, Research & Evaluation*, 6(2). Available online: http://ericae.net/pare/getvn.asp?v=6&n=2. This paper has been viewed 34783 times since 11/13/99.

Implementing Performance Assessment in the Classroom

Amy Brualdi,
ERIC/AE

Introduction

If you are like most teachers, it probably is a common practice for you to devise some sort of test to determine whether a previously taught concept has been learned before introducing something new to your students. Probably, this will be either a completion or multiple choice test. However, it is difficult to write completion or multiple choice tests that go beyond the recall level. For example, the results of an English test may indicate that a student knows each story has a beginning, a middle, and an end. However, these results do not guarantee that a student will write a story with a clear beginning, middle, and end. Because of this, educators have advocated the use of performance-based assessments.

Performance-based assessments "represent a set of strategies for the . . . application of knowledge, skills, and work habits through the performance of tasks that are meaningful and engaging to students" (Hibbard and others, 1996, p. 5). This type of assessment provides teachers with information about how a child understands and applies knowledge. Also, teachers can integrate performance-based assessments into the instructional process to provide additional learning experiences for students.

The benefit of performance-based assessments are well documented. However, some teachers are hesitant to implement them in their classrooms. Commonly, this is because these teachers feel they don't know enough about how to fairly assess a student's performance (Airasian,1991). Another reason for reluctance in using performance-based assessments may be previous experiences with them when the execution was unsuccessful or the results were

inconclusive (Stiggins, 1994). The purpose of this digest is to outline the basic steps that you can take to plan and execute effective performance-based assessments.

Defining the Purpose of the Performance-Based Assessment

In order to administer any good assessment, you must have a clearly defined purpose. Thus, you must ask yourself several important questions:

- What concept, skill, or knowledge am I trying to assess?
- What should my students know?
- At what level should my students be performing?
- What type of knowledge is being assessed: reasoning, memory, or process (Stiggins, 1994)?

By answering these questions, you can decide what type of activity best suits you assessment needs.

Choosing the Activity

After you define the purpose of the assessment, you can make decisions concerning the activity. There are some things that you must take into account before you choose the activity: time constraints, availability of resources in the classroom, and how much data is necessary in order to make an informed decision about the quality of a student's performance (This consideration is frequently referred to as sampling.).

The literature distinguishes between two types of performance-based assessment activities that you can implement in your classroom: informal and formal (Airasian, 1991; Popham, 1995; Stiggins, 1994). When a student is being informally assessed, the student does not know that the assessment is taking place. As a teacher, you probably use informal performance assessments all the time. One example of something that you may assess in this manner is how children interact with other children (Stiggins, 1994). You also may use informal assessment to assess a student's typical behavior or work habits.

A student who is being formally assessed knows that you are evaluating him/her. When a student's performance is formally assessed, you may either have the student perform a task or complete a project. You can either observe the student as he/she performs specific tasks or evaluate the quality of finished products.

You must beware that not all hands-on activities can be used as performance-based assessments (Wiggins, 1993). Performance-based assessments require individuals to apply their knowledge and skills in context, not merely completing a task on cue.

Defining the Criteria

After you have determined the activity as well as what tasks will be included in the activity, you need to define which elements of the project/task you shall to determine the success of the student's performance. Sometimes, you may be able to find these criteria in local and state curriculums or other published documents (Airasian, 1991). Although these resources may prove to be very useful to you, please note that some lists of criteria may include too many skills or concepts or may not fit your needs exactly. With this in mind, you must be certain to review criteria lists before applying any of them to your performance-based assessment.

You must develop your own criteria most of the time. When you need to do this, Airasian (1991, p. 244) suggests that you complete the following steps:

1. Identify the overall performance or task to be assessed, and perform it yourself or imagine yourself

performing it
2. List the important aspects of the performance or product.
3. Try to limit the number of performance criteria, so they can all be observed during a pupil's performance.
4. If possible, have groups of teachers think through the important behaviors included in a task.
5. Express the performance criteria in terms of observable pupil behaviors or product characteristics.
6. Don't use ambiguous words that cloud the meaning of the performance criteria.
7. Arrange the performance criteria in the order in which they are likely to be observed.

You may even wish to allow your students to participate in this process. You can do this by asking the students to name the elements of the project/task that they would use to determine how successfully it was completed (Stix, 1997).

Having clearly defined criteria will make it easier for you to remain objective during the assessment. The reason for this is the fact that you will know exactly which skills and/or concepts that you are supposed to be assessing. If your students were not already involved in the process of determining the criteria, you will usually want to share them with your students. This will help students know exactly what is expected of them.

Creating Performance Rubrics

As opposed to most traditional forms of testing, performance-based assessments don't have clear-cut right or wrong answers. Rather, there are degrees to which a person is successful or unsuccessful. Thus, you need to evaluate the performance in a way that will allow you take those varying degrees into consideration. This can be accomplished by creating rubrics.

A rubric is a rating system by which teachers can determine at what level of proficiency a student is able to perform a task or display knowledge of a concept. With rubrics, you can define the different levels of proficiency for each criterion. Like the process of developing criteria, you can either utilize previously developed rubrics or create your own. When using any type of rubric, you need to be certain that the rubrics are fair and simple. Also, the performance at each level must be clearly defined and accurately reflect its corresponding criterion (or subcategory) (Airasian, 1991; Popham, 1995; Stiggins, 1994).

When deciding how to communicate the varying levels of proficiency, you may wish to use impartial words instead of numerical or letter grades (Stix, 1997). For instance, you may want to use the following scale: word, sentence, page, chapter, book. However, words such as "novice," "apprentice," "proficient," and "excellent" are frequently used.

As with criteria development, allowing your students to assist in the creation of rubrics may be a good learning experience for them. You can engage students in this process by showing them examples of the same task performed/project completed at different levels and discuss to what degree the different elements of the criteria were displayed. However, if your students do not help to create the different rubrics, you will probably want to share those rubrics with your students before they complete the task or project.

Assessing the Performance

Using this information, you can give feedback on a student's performance either in the form of a narrative report or a grade. There are several different ways to record the results of performance-based assessments (Airasian,1991; Stiggins,1994):

- *Checklist Approach* When you use this, you only have to indicate whether or not certain elements are present in the performances.

- *Narrative/Anecdotal Approach* When teachers use this, they will write narrative reports of what was done during each of the performances. From these reports, teachers can determine how well their students met their standards.
- *Rating Scale Approach* When teachers use this, they indicate to what degree the standards were met. Usually, teachers will use a numerical scale. For instance, one teacher may rate each criterion on a scale of one to five with one meaning "skill barely present" and five meaning "skill extremely well executed."
- *Memory Approach* When teachers use this, they observe the students performing the tasks without taking any notes. They use the information from their memory to determine whether or not the students were successful. (Please note that this approach is not recommended.)

While it is a standard procedure for teachers to assess students' performances, teachers may wish to allow students to assess them themselves. Permitting students to do this provides them with the opportunity to reflect upon the quality of their work and learn from their successes and failures.

References and Additional Reading

Airasian, P.W. (1991). *Classroom assessment.* New York : McGraw-Hill.

Hibbard, K. M. and others. (1996). *A teacher's guide to performance-based learning and assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Popham, W. J. (1995). *Classroom assessment: What teachers need to know.* Needham Heights, MA: Allyn and Bacon.

Stiggins, R. J. (1994). *Student-centered classroom assessment.* New York: Macmillan Publishing Company.

Stix, A. (1997). Empowering students through negotiable contracting. (Paper presented at the National Middle School Initiative Conference (Long Island, NY, January 25, 1997) (ERIC Document Reproduction Number ED411274)

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, May, 703-713.

Wiggins, G. (1993). Assessment, authenticity, context, and validity. *Phi Delta Kappan,* November, 200-214.

Wiggins,G. (1998). *Educative assessment: designing assessments to inform and improve student performance* San Francisco, Calif. : Jossey-Bass.

Descriptors: *Criteria; *Educational Assessment; *Evaluation Methods; Feedback; *Performance Based Assessment; Sampling; Scoring; Self Evaluation (Individuals); *Student Evaluation; Teaching Methods

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

Shadish, William (1998). Some evaluation questions. *Practical Assessment, Research & Evaluation*, 6(3). Available online: http://ericae.net/pare/getvn.asp?v=6&n=3. This paper has been viewed 25364 times since 11/13/99.

Some Evaluation Questions

William Shadish
Memphis State University

This article was adapted from Shadish, W. (1998) Presidential Address: Evaluation Theory is Who We Are. *American Journal of Evaluation,* 19, 1, 1-19.

Introduction

In question and answer format, this digest illustrates the variety of basic and theoretical issues in evaluation with which aspiring evaluators should be conversant in order to claim they know the knowledge base of their profession. Please note that none of these questions have a single correct answer and space limitations prevent providing the level of detailed discussion that each deserves. The questions vary considerably in difficulty and in how universally the issues involved would be recognized by most evaluators today. What follows, therefore, are outlines of the issues rather than correct "answers." For more extensive information on the topics discussed in these questions, please refer to the references found at the end of this digest, especially Shadish, Cook, and Leviton (1991).

What are the four steps in the logic of evaluation?

Scriven (1969, 1989) published a variety of writings on the topic of the logical sequence of concepts that defines how people try to connect data to value judgments that the evaluand is good or bad, better or worse, passing or failing, or the like. Scriven outlined the four steps in 1980:

- selecting criteria of merit, those things the evaluand must do to be judged good
- setting standards of performance on those criteria, comparative or absolute levels that must be exceeded to warrant the appelation "good"
- gathering data pertaining to the evaluand's performance on the criteria relative to the standards

- integrating the results into a final value judgment.

To the extent that evaluation really is about determining value, some version of this logic ought to be universally applicable to the practice of evaluation.

Are qualitative evaluations valid?

More qualitative theorists than not seem to both use the word, subject to validity criticism, and endorse some version of its applicability. However, some qualitative theorists reject both the term and any cognates who seem to garner attention disproportionate to their representation in their own field. From outside the qualitative camps, the answer also seems to be more uniformly "yes." Nevertheless, the subtleties required for an intelligent discussion of this question are extensive, of which the following few will illustrate but not exhaust. Even those who reject the concept of "validity" will acknowledge they are concerned in their work to "go to considerable pains not to get it all wrong" (Wolcott,1990, p. 127). Further, within and across those methods qualitative theorists often disagree among themselves. In addition, qualitative methods often aim to produce knowledge of a substantively different kind than other methods, so that particular validity criteria may be less pertinent to the interests of qualitative evaluations. Indeed, it would be wrong to assume all qualitative methods are alike. Different qualitative methods may have different aims that bring different validity criteria to bear. In the end, though, some version of validity as an effort to "go to considerable pains not to get it all wrong"(Wolcott, 1990, p. 127) probably underlies all methods used by all evaluators, quantitative and qualitative.

What difference does it make whether the program being evaluated is new or has existed for many years?

Rossi and Freeman (e.g., 1993) long made this distinction central to their approach to evaluation because it has several implications for evaluation practice. Brand-new programs have not yet had time to work out program conceptualization and implementation problems. Thus, a focus on those kinds of questions is likely to be more useful and more acceptable to program staff than a focus on, say, outcome questions. In addition, less background information and fewer past evaluations are likely to exist for new programs, so more work will have to be done "from scratch." Well-established programs may be more ready for outcome evaluation, and they may have a greater wealth of information already available on them. However, long-established programs may also have reached so many of the potential participants that outcome evaluations might be thwarted by difficulty finding appropriate control group participants if a controlled design is used.

Is there a difference between evaluating a large program, a local project within that program, or a small element within that project?

This distinction points to an interesting tradeoff between ease and frequency of short-term change on the one hand, and likely impact on the other (Cook, Leviton, & Shadish, 1985; Shadish, Cook, & Leviton, 1991). Small elements have natural turnover rates that are much more frequent than for local projects, which themselves turnover less often than large programs. Hence, the opportunity to change each of them by replacement is more frequent for smaller than larger entities. However, smaller entities are usually likely to have a smaller impact on the overall set of problems to which the program, project, or elements are aimed. All this has implications for the kinds of questions worth asking depending on what kind of use and impact is desired.

How can the chances of evaluation results being used in the short-term to make changes be increased?

The literature on this topic is extensive (e.g., Cousins & Shuhla, 1997; Patton, 1997), and includes advice to locate a powerful user(s), identify questions of interest to the user(s), focus on things that the user has sufficient control over to change, discuss exactly what changes the user(s) would make given different kinds of answers that might result from the evaluation, provide interim findings at points when they might be useful, consider reporting results in both

traditional and nontraditional formats, provide brief executive summaries of results, have continued personal contact after the evaluation ends, and lend support to subsequent efforts to foster use of evaluation results.

What are the disadvantages of focusing on short-term instrumental use?

There is a risk that the evaluation will focus on less important interventions or questions than might otherwise be the case, and lose the big picture or the long-term outlook about what is important. In part this reflects the tradeoffs discussed regarding the program-project-element distinction because instrumental use is more likely with smaller elements likely to have less impact. It also reflects the fact that the modern industrial societies where much evaluation takes place have often solved the easiest problems, so those that remain are often difficult to do anything about in the short-term. Those things that can be addressed in the short-term are rarely likely to fall into the set of most difficult problems. Finally, it is rare to find a user who can control options that promise truly powerful or fundamental changes.

What role does causal inference play in evaluation?

The most obvious version of this question concerns the role of outcome evaluation. From an early dependency on outcome evaluation as paradigmatic for the field, the field realized the value of asking a wide array of other questions depending on contingencies like those discussed previously regarding use, program size, and stage of program development So causal inference of the traditional sort assumed a smaller role in evaluation than in early years. Another version of this question appeals to the distinction between descriptive causal inferences and causal mediation; the latter has enjoyed some recent resurgence in some kinds of theory-driven evaluation.

Would the answer change if questions were asked about the role that causal inference played in making value judgements?

Most readers probably assumed "evaluation" in the previous question to mean the wide range of activities that fall under the rubric of professional evaluation practice. This question plays on limiting the meaning of the term "evaluation" to the activity of making a value judgment. Some readers might not realize that, even in this limited context, causal inference still plays an important role. Referring back to the answer to the first question about the logic of evaluation, implicit in it in most applications is that the thing being evaluated caused the observed performance on the criteria of merit (e.g., that the treatment met recipient needs). If that were not the case, it would be improper to attribute the merit or value to the evaluand; rather, it should be attributed to whatever else actually caused the improvement in the criteria of merit. Thus attributing merit or worth is frequently causal in substantial part.

When does a question have leverage?

Cronbach and his colleagues (Cronbach, Ambron, Dornbusch, Hess, Hornik, Phillips, Walker, & Weiner, 1980) used this term to describe questions they thought particularly worth asking because of their potential for high payoff. Such questions have little prior information available, they can be feasibly answered with the resources and in the time available, the answers will probably reduce uncertainty significantly, and the answers are of interest to the policy shaping community.

What is metaevaluation, and when should it be used?

Metaevaluation is the evaluation of evaluation (Cook & Gruder, 1978; Scriven, 1969), and recommendations vary from doing it for every evaluation to doing in periodically. The general prescription is that metaevaluation can be done using the same general logic (and sometimes methods) for doing the primary evaluation. One might apply the logic of evaluation from the first question, for example, asking what the evaluation would do well to be a good

evaluation (e.g., would it be useful, true, important?), deciding how well would it do so (e.g., how useful? true by what standards?),measuring the performance of the evaluation in these regards, and then synthesizing results to reach a judgment about the merits of the evaluation. Metaevaluation can be applied at nearly any stage of an evaluation, from evaluating its planned questions and methods, to a mid-evaluation review, to evaluating the completed report by submitting it to independent consultants and critics.

References and Additional Reading

Campbell, D. T. (1971). *Methods for experimenting society.* Paper presented at the meeting of the Eastern Psychological Association, New York and at the meeting of the American Psychological Association, Washington, DC.

Cook, T. D. and Gruder, C. L. (1978). Metaevaluative research. *Evaluation Quarterly, 2*, 5-51.

Cook, T. D., Leviton, L. C., and Shadish, W. R. (1985). Program Evaluation. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (3rd ed., pp. 699-777). New York: Random House.

Cronach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D.C., Walker, D. F., and Weiner, S. S. (1980). *Toward Reform of Program Evaluation.* San Francisco: Jossey Bass.

Lincoln, Y. S. (1990). Program Review, Accreditation processes, and outcomes assessment: Pressures on institutions of higher education. *Evaluation Practice, 11*(1), 13-25.

Maxwell, J. A. (1992). Understanding validity in qualitative research. *Harvard Educational Review, 62*(3), 279-300.

Patton, M. Q. (1997). *Utilization-focused Evaluation: the New Century Text.* Thousand Oaks, CA: Sage Publications.

Rossi, P. H. & Freeman, H. E. (1993). *Evaluation: a Systemic Approach.* Newbury Park, CA: Sage Publications.

Scriven, M. (1969). An introduction to meta-evaluation. *Educational Product Report, 2,* 36-38.

Scriven, M. (1980). *The Evaluation of College Teaching.* Syracuse, N.Y.: National Dissemination Center, Syracuse University School of Education.

Shadish, W. R., Cook, T. D., & Leviton, L. D. (1991). *Foundations of Program Evaluation: Theories of Practice.* Newbury Park, CA: Sage Publications.

Descriptors: Causal Models; Criteria; *Evaluation Methods; Evaluation Problems; Evaluation Utilization; *Evaluators; *Inferences; Performance Factors; *Program Evaluation; Qualitative Research; Standards; *Theory Practice Relationship; User Needs (Information)

Home | Articles | Subscribe | Review | Policies

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

Search: title Go

Rudner, Lawrence (1998). Item banking. *Practical Assessment, Research & Evaluation*, 6(4). Available online: http://ericae.net/pare/getvn.asp?v=6&n=4. This paper has been viewed 6826 times since 11/13/99.

Item Banking

Lawrence Rudner,
ERIC Clearinghouse on Assessment and Evaluation

Various school districts use standardized tests as a way to measure scholastic achievement. Usually, these districts need to revise tests with some frequency to avoid administering the same test year after year. Unfortunately, creating new tests can be a very time consuming endeavor. Not only do test writers need to compose the test items, they also must determine each item's difficulty in order to ensure that a test will neither be too hard nor too easy.

Using item banks, test makers can escape this process. Item banks are files of various suitable test items that are "coded by subject area, instructional level, instructional objective measured, and various pertinent item characteristics (e.g., item difficulty and discriminating power)" (Gronlund, 1998, p. 130). The purpose of this digest is to discuss the advantages and disadvantages of using item banks as well as provide useful information to those who are considering implementing an item banking project in their school district.

Advantages of Item Banking

The primary advantage of item banking is in test development. Using a item response theory method, such as the Rasch model, items from multiple tests are placed on a common scale, one scale per subject matter. The scale indicates the relative difficulty of the items. Items can be placed on the scale, i.e. into the item bank, without extensive testing. New subtests and tests, with predictable characteristics, can be developed by drawing items from the bank. For example, suppose you are interested in developing a new subtest to cover fractions in seventh grade. You can go to the item bank, identify items related to your objectives and then predict the characteristics of a subtest composed of those items. The effect of including or excluding particular items can also be predicted.

Another advantage of an item bank is that it will permit you to "deposit" additional items to be withdrawn as

needed. Depending on the size of the testing program, there can be two practical approaches for making deposits. You can make "large deposits" by merging your item bank with one from another district. You can also make "small deposits" by adding a few locally developed items at a time. The large deposit option will involve purchasing or trading items with another district and then equating their scale to yours. The small deposit option involves piloting a fewer number of items with examinees in several grade levels. This can easily be accomplished by adding a supplemental page containing experimental items to be administered along booklet from the school system.

Item banking provides substantial savings of time and energy over conventional test development. In traditional test development, items can only be described relative to the other items within the test and to whom they were given. That is, item characteristics are extremely group and test specific. With item banking, items are described their relative difficulty across grade levels. In order to develop a new test or subtest, one does not need to go through the laborious process of developing a large set of items for piloting and evaluating. Instead, one just draws from the bank. Further, drawing from the bank allows one to make fairly accurate predictions concerning composite test characteristics.

One additional advantage of item banking is that it helps establish a language for discussing curriculum goals and objectives. The items describe individual tasks students are capable or incapable of doing. The location of the items on a calibrated scale allows one to identify the relative difficulty of particular tasks. This provides a way to discuss possible learning hierarchies and ways to better structure curriculum.

Disadvantages and Limitations of Item Banking

Item banking and item response theory are not cure-alls for measurement problems. Persistence and good judgement must remain vital aspects in any test construction and test usage effort. One must make every possible effort to include only quality items in the item bank. The same care and effort must go into item writing. Items purchased form external sources must be evaluated carefully for match to your curriculum as well as for technical quality.

Item banking involves equating various tests and items. It is entirely possible, mathematically, to equate tests which cover entirely different subject matter. At the practical level, this means that it is also possible to equate items which assess subtly, but significantly different skills. In order to avoid this undesirable situation, the item review process must also include a careful evaluation of the skills assessed by each item and tests must be carefully formulated.

The intent of compiling a test using latent trait theory is to be able to make a prediction of the composite test characteristics. While the prediction is often surprisingly accurate, it must be validated. Tests developed using latent trait theory should still be field tested.

While some districts have implemented very successful item banks and Rasch calibrated testing programs without knowing anything about IRT, good practice calls for a staff that is comfortable with and knowledgeable of what they are doing. A district undertaking an item banking project should have full understanding of the practical as well as the mathematical/theoretical aspects of item banking.

An item bank really consists of multiple collections of items with fairly unidimensional content area, such as mathematic computations or vocabulary. Collections of items usually span several grade levels. In order to develop the bank, many tests must be calibrated, linked (or equated), and organized. This requires a great deal of work in terms of preparation and planning and in terms of computer time and expertise. Once the item bank is established, however, test development time, effort, and cost is reduced.

Planning for an Item Bank

The most crucial step in developing an item bank is planning. This involves the preparation of individuals, the identification of what you have to start an item bank, and the identification of what you hope to accomplish with an item bank.

Everyone on the staff should have enough familiarity with Rasch measurement principles and item banking to be able to knowledgeably discuss and explain the project. You can formally train your staff by using in-house personnel, bringing in a traveling workshop, or having people attend a pre-session at a research association or conference.

You should have senior level personnel available to answer technical questions that might arise. You should also have computer experts that are capable of doing the following tasks: 1.) modifying computer programs, 2.) establishing a data base system, and 3.) capable of running packaged programs.

If you intend to do any item bank exchanges or purchases, you should have someone on your staff who knows what is available. You need personnel capable of critically evaluating test items for technical quality, curriculum match, unidimensionality, and potential bias. In order to accurately calibrate test items and establish scales, items need to be presented to examinees with a wide range of ability.

In order to link various forms and grade levels within a content area, common anchor items are needed. (These anchor items must be administered along with the items within a given form. The form and anchor items are calibrated together. The anchor item parameter values based on calibration with one form are compared with the anchor item parameter values based on calibration with another form. The difference in parameter values is used to link the forms.) You need to identify for which content areas you have administered overlapping subtests and the number of students responding to the set of items. You may find you will need to gather additional item response data to link forms and grade levels.

Your data processing staff should examine literature and programs on item banking to determine what programs must be developed and what programs can be modified.

As much as possible, you should identify your projected testing needs for the next five years. This would involve identification of which subtests you will need to revise, what additional areas you may need to assess, and how objectives might be differently stressed.

Start-up Activities

The start-up activities would mostly involve administrative activities and the data processing staff. Each test would have to be calibrated and equated to the parallel form and adjacent grade levels. The data processing staff would have to adapt existing computer programs to the local system and develop a database system. They would then calibrate each test, equate the tests, and store the equated item parameters and their descriptors in a database system. With a large number of tests and items, this becomes a major undertaking.

Administrative staff would have to coordinate activities to insure that the data requirements are met. During the planning process, a chart can be developed to identify which tests and anchor items have been and will need to be administered to the requisite sample. Working from these charts, testing coordinators will need to organize the administration of tests and subtests needed to calibrate and equate all the items going into the item bank. This involves compiling test booklets, making testing arrangements, collecting response sheets, and preparing data for data processing. Depending on frequency of students taking multiple subtests from different levels and forms, this too can be a major undertaking.

Running the Item Bank

The item bank will allow you to withdraw items as needed to develop new or even special tests and subtests. There are basically two activities involved in running an item bank - making deposits and withdrawing items to develop a test.

As mentioned earlier, there are to viable options for making deposits to the item bank. The "large deposit" option involves merging an existing item bank with your own. If the existing item bank has been IRT calibrated, then you only need to administer a subset of items (per content area) from the new bank along with items already in your item bank. Remember, each item bank uses its own anchor items and allows you to equate the scales. This part will involve testing with a relatively small group of students. The anchor items from the new item bank can be appended to present group. Coordination would be similar to that involved in starting your own item bank.

The major task involved in using items from another item bank is a thorough, careful review of the items. All potential entries must be evaluated for technical quality, curriculum match, and potential bias. This would involve your test development experts, curriculum/instructional staff, and coordination between the two.

After an item review, items from non-calibratcd could be treated like items developed by your staff. "Small deposits" would be made by calibrating and equating a few items at a time. One very efficient approach to collecting the requisite data is to append subtests of new items to original groups. The items within the original group would serve as anchor items for the new subtest(s) of items. In this manner, you can be constantly adding to your item bank.

Once developed and growing, your item bank is ready to provide the advantages discussed above. To develop a new subtest, you would develop a blueprint/table of specifications to outline what you want your new subtest to be like. Curriculum specialists and test development experts would then go to the item bank and identify which item in the bank appear appropriate in terms of content and in terms of their relative difficulty. If they find an insufficient number of items, them can make arrangements to add new items to the bank.

If the bank contains a sufficient number of items of the appropriate nature, the items can be grouped to form a new subtest. Without pilot testing, the characteristics of this new subtest can be predicted. With reasonable accuracy, you will know how much skill an examinee needs to obtain any given total raw score on the new subtest. The prediction should be validated by administering the subtest to students having received appropriate instruction and students not having received such instruction. This can also be accomplished by appending items to the existing forms. This validation would need a sample as large as you used in field testing the original group.

An item bank provides a scale of relative difficulty of tasks that covers multiple grade levels and skills within content areas. As a service to the instructional/curriculum staff, you can provide information on the relative difficulty of different taks within and across grades levels. For example, you can identify which fraction problems seventh graders find as difficult as certain decimal problems; or you can identify which reading skills taught in fourth grade can be mastered by students in their grade. It could also be used to help organize special programs for gifted and remedial students.

Additional Reading

Grolund, N.E. (1998). *Assessment of Student Achievement. Sixth Edition.* Needham Heights, MA: Allyn and Bacon.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, N.J. : L. Erlbaum

Associates.

Mengel, Bill E.; Schorr, Larry L. (1992) Developing Item Bank Based Achievement Tests and Curriculum-Based Measures: Lessons Learned Enroute. (ERIC Document Reproduction Number ED344915).

Ward, A.W.; Murray-Ward, M. (1994). Guidelines for the development of item banks. An NCME instructional module. *Educational Measurement: Issues and Practice,13*(1), 34-39.

Wright, B.D.; Stone, M.H. (1979). *Best Test Design. Rasch Measurement.* Chicago, IL: MESA Press.

Descriptors: Adaptive Testing; *Computer Assisted Testing; Difficulty Level; *Item Banks; Item Response Theory; *Test Construction; Test Items

Home | Articles | Subscribe | Review | Policies

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

Search: title Go

Brualdi, Amy (1998). Teacher comments on report cards. *Practical Assessment, Research & Evaluation*, 6(5). Available online: http://ericae.net/pare/getvn.asp?v=6&n=5.
This paper has been viewed 16349 times since 11/13/99.

Teacher Comments on Report Cards

by Amy Brualdi
ERIC Clearinghouse on Assessment and Evaluation

Introduction

Several times a year, teachers must complete a report card for each student in order to inform parents about the academic performance and social growth of their child. Schools have a variety of ways to document the progress of students. In a majority of schools, teachers usually assign a number or letter grade to the subject or skill areas. In several schools, mostly elementary schools, teachers write a descriptive narrative of each child's cognitive and social growth. Other schools have teachers indicate whether a student has acquired different skills by completing a checklist.

Despite the fact that schools have different policies concerning the report card's content and format, most teachers are required to include written comments about the student's progress. Considering the amount of students in each classroom, the long span of time needed to complete each report card, and the presence of grade/check marks on the report cards, some may think that comments are nonessential and take up too much of a teacher's time. The purpose of this digest is to explain why teacher comments on report cards are important, offer suggestions on how to construct effective comments, point out words or phrases to be cautious of using, and indicate sources of information for report card comments.

Why are comments important?

Grades are designed to define the student's progress and provide information about the skills that he/she has or has not acquired. Nevertheless, grades are often not detailed enough to give parents or the student him/herself a thorough understanding of what the he/she has actually learned or accomplished (Wiggins, 1994; Hall, 1990). For example, if a child receives a B in spelling, a report card comment can inform the parent that the child is generally a

good speller; however, she consistently forgets to add an es to plural nouns ending with the letters, s and x. Thus, teacher comments often convey whatever information has not been completely explained by the grade.

Well written comments can give parents and children guidance on how to make improvements specific academic or social areas. For example, the teacher who wrote the previous report card comment on spelling may also wish to include that practicing how to write the different plural nouns at home or playing different spelling games may help the child to enhance her spelling skills.

The process of writing comments can also be helpful to teachers. Writing comments gives teachers opportunities to be reflective about the academic and social progress of their students. This time of reflection may result in teachers gaining a deeper understanding of each student's strengths and needs.

What types of wording should teachers include in their comments?

The use of specific comments encourages positive communication between teachers, parents, and students. Written in a positive and informative manner, comments can address a variety of issues while maintaining the while still maintaining the dignity of the child. This is especially important if a child has had difficulty with a particular subject area or controlling his/her behavior over an extended period of time.

Shafer (1997) compiled a list of "effective" comments from a variety of teachers. The following lists of words and phrases are just a sampling from her publication "Writing Effective Report Card Comments" (p. 42-43).

Words that promote positive view of the student

- thorough
- caring
- shows commitment
- improved tremendously
- has a good grasp of

Words and Phrases to use to convey that a child needs help

- could profit by
- requires
- finds it difficult at times to
- needs reinforcement in
- has trouble with

Words and phrases that teachers should be cautious of using

When teachers write comments on report cards, they need to be cognizant of the fact that each child has a different rate of social and academic development. Therefore, comments should not portray a child's ability as fixed and permanent (Shafer, 1997). Such comments do not offer any reason to believe that the child will be successful if he/she attempts to improve.

Also, teachers must be sensitive to the fact that their students will read their comments. If negative comments are made, teachers must be aware that those comments may be counterproductive. In addition to the previously mentioned positive comments, Shafer (1997) compiled a list of words and phrases that should be avoided or used with caution (p. 45).

Words to Avoid or Use with Caution

- unable
- can't
- won't
- always
- never

Information sources to which teachers should look when writing report card comments

Teachers should have a plethora of sources from which they can derive information on each child to support the comments that are made on each report card. Teachers need these in order to provide specific information on the different strengths and weaknesses of each child. The most commonly used sources of information are examples of student work and test results. In addition to these traditional sources, teachers also use student portfolios as well as formal and informal student observations.

Arter, Spandel, and Culham (1995) define the student portfolio as "a purposeful collection of student work that tells the story of student achievement and growth" (p. 1). A student's portfolio is usually comprised of work that is either the student's best or most exemplary of his/her ability. A portfolio may also contain papers which show the evolution of a particular writing assignment or project. In addition to aiding teachers in keeping track of a student's progress, the portfolio allows the student to chart his/her own academic growth. Because of this, a student should not have many surprises on his report card and will understand how he earned his his grades and why different teacher comments were written.

Another rich source of information is the student observation. Student observations often provide important information that is sometimes difficult to derive from the written work of students. These observations allow teachers to make comments on students' daily academic and social behaviors. These should be written about the students' behaviors in a variety of settings: independent work, cooperative learning groups, and playground or nonacademic interaction (Grace, 1992). Grace (1992) suggests that teachers have the following observations for each child: anecdotal records, checklist or inventory, rating scales, questions and requests, and results from screening tests.

References and Additional Readings

Arter J. A., Spandel, V., Culham, R. (1995). Portfolios for assessment and instruction. (ERIC Document Reproduction Service ED388890).

Farr, R. (1991). Portfolios: Assessment in language arts. ERIC digest. (ERIC Document Reproduction Service ED334603).

Grace, C. (1992). The portfolio and its use: Developmentally appropriate assessment of young children. ERIC digest. (ERIC Document Reproduction Service ED351150).

Guskey, T.R. (Ed.) (1996). . . *Association of Supervision and Curriculum Development Yearbook 1996. Communicating Student Progress.* Arlington, VA: ASCD.

Guskey, T.R. (1996). Reporting on student learning: Lessons from the past- Prescriptions for the future . *Association of Supervision and Curriculum Development Yearbook 1996. Communicating Student Progress.* Arlington, VA: ASCD, pp. 13-24.

Hall, K. (1990). Determining the success of narrative report cards. (ERIC Document Reproduction Service No. 334 013).

Lake, K. and Kafka, K. (1996). Reporting methods in grades K-8. *Association of Supervision and Curriculum Development Yearbook 1996. Communicating Student Progress.* Arlington, VA: ASCD. pp. 90-118

Peckron, K.B. (1996). Beyond the A: Communicating the learning progress of gifted students. . *Association of Supervision and Curriculum Development Yearbook 1996. Communicating Student Progress.* Arlington, VA: ASCD pp. 58-64.

Shafer, S. (1997). *Writing Effective Report Card Comments.* New York, NY: Scholastic. [Amazon]

Wiggins, G. (1994). Toward better report cards. *Educational Leadership. 52*(2). pp. 28-37

Descriptors: *Academic Achievement; *Communication (Thought Transfer); Elementary Secondary Education; Evaluation Methods; *Parent Teacher Cooperation; *Report Cards; *Student Evaluation; *Teacher Attitudes

Home | Articles | Subscribe | Review | Policies

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

Search: title Go

Brualdi, Amy C. (1998). Classroom questions. *Practical Assessment, Research & Evaluation*, 6(6). Available online: http://ericae.net/pare/getvn.asp?v=6&n=6. This paper has been viewed 17232 times since 11/13/99.

Classroom Questions

Amy C. Brualdi,
ERIC Clearinghouse on Assessment and Evaluation

In 1912, Stevens stated that approximately eighty percent of a teacher's school day was spent asking questions to students. More contemporary research on teacher questioning behaviors and patterns indicate that this has not changed. Teachers today ask between 300-400 questions each day (Leven and Long, 1981).

Teachers ask questions for several reasons (from Morgan and Saxton, 1991):

- the act of asking questions helps teachers keep students actively involved in lessons;
- while answering questions, students have the opportunity to openly express their ideas and thoughts;
- questioning students enables other students to hear different explanations of the material by their peers;
- asking questions helps teachers to pace their lessons and moderate student behavior; and
- questioning students helps teachers to evaluate student learning and revise their lessons as necessary.

As one may deduce, questioning is one of the most popular modes of teaching. For thousands of years, teachers have known that it is possible to transfer factual knowledge and conceptual understanding through the process of asking questions. Unfortunately, although the act of asking questions has the potential to greatly facilitate the learning process, it also has the capacity to turn a child off to learning if done incorrectly. The purpose of this digest is to provide teachers with information on what types of question and questioning behaviors can facilitate the learning process as well as what types of questions are ineffective.

What is a Good Question?

In order to teach well, it is widely believed that one must be able to question well. Asking good questions fosters interaction between the teacher and his/her students. Rosenshine (1971) found that large amounts of student-teacher

interaction promotes student achievement. Thus, one can surmise that good questions fosters student understanding. However, it is important to know that not all questions achieve this.

Teachers spend most of their time asking low-level cognitive questions (Wilen, 1991). These questions concentrate on factual information that can be memorized (ex. What year did the Civil War begin? or Who wrote *Great Expectations*?). It is widely believed that this type of question can limit students by not helping them to acquire a deep, elaborate understanding of the subject matter.

High-level-cognitive questions can be defined as questions that requires students to use higher order thinking or reasoning skills. By using these skills, students do not remember only factual knowledge. Instead, they use their knowledge to problem solve, to analyze, and to evaluate. It is popularly believed that this type of question reveals the most about whether or not a student has truly grasped a concept. This is because a student needs to have a deep understanding of the topic in order to answer this type of question. Teachers do not use high-level-cognitive questions with the same amount of frequency as they do with low-level-cognitive questions. Ellis (1993) claims that many teachers do rely on low-level cognitive questions in order to avoid a slow-paced lesson, keep the attention of the students, and maintain control of the classroom.

Arends (1994) argues that many of the findings concerning the effects of using lower-level-cognitive versus higher-level-cognitive questions has been inconclusive. While some studies and popular belief favor asking high-level-cognitive, other studies reveal the positive effects of asking low-level cognitive questions. Gall (1984), for example, cited that "emphasis on fact questions is more effective for promoting young disadvantaged children's achievement, which primarily involves mastery of basic skills; and emphasis on higher cognitive questions is more effective for students of average and high ability. . ." (p. 41). Nevertheless, other studies do not reveal any difference in achievement between students whose teachers use mostly high level questions and those whose teachers ask mainly low level questions (Arends, 1994; Wilen, 1991). Therefore, although teachers should ask a combination of low-level-cognitive and high-level-cognitive questions, they must determine the needs of their students in order to know which sort of balance between the two types of questions needs to be made in order to foster student understanding and achievement.

How to ask questions that foster student achievement

In a research review on questioning techniques, Wilen and Clegg (1986) suggest teachers employ the following research supported practices to foster higher student achievement:

- phrase questions clearly;
- ask questions of primarily an academic nature
- allow three to five seconds of wait time after asking a question before requesting a student's response, particularly when high-cognitive level questions are asked;
- encourage students to respond in some way to each question asked;
- balance responses from volunteering and nonvolunteering students;
- elicit a high percentage of correct responses from students and assist with incorrect responses;
- probe students' responses to have them clarify ideas, support a point of view, or extend their thinking;
- acknowledge correct responses from students and use praise specifically and discriminately. (p. 23)
-

What is a Bad Question?

When children are hesitant to admit that they do not understand a concept, teachers often try to encourage them to ask questions by assuring them that their questions will neither be stupid or bad. Teachers frequently say that all questions have some merit and can contribute to the collective understanding of the class. However, the same theory

does not apply to teachers. The content of the questions and the manner in which teachers ask them determines whether or not they are effective. Some mistakes that teachers make during the question and answer process include the following: asking vague questions (ex. What did you think of the story that we just read?), asking trick questions, and asking questions that may be too abstract for children of their age (ex. asking a kindergarten class the following question: How can it be 1:00 P.M. in Connecticut but 6:00 P.M. in the United Kingdom at the same moment?)

When questions such as those mentioned are asked, students will usually not know how to respond and may answer the questions incorrectly. Thus, their feelings of failure may cause them to be more hesitant to participate in class (Chuska, 1995), evoke some negative attitudes towards learning, and hinder the creation of a supportive classroom environment.

Conclusion

Sanders (1966) stated, "Good questions recognize the wide possibilities of thought and are built around varying forms of thinking. Good questions are directed toward learning and evaluative thinking rather than determining what has been learned in a narrow sense" (p. ix). With this in mind, teachers must be sure that they have a clear purpose for their questions rather than just determining what knowledge is known. This type of question planning results in designing questions that can expand student's knowledge and encourage them to think creatively.

References and Additional Readings

Arends, R. (1994). *Learning to teach.* New York, NY: McGraw-Hill, Inc.

Bloom, B., Englehart, M., Furst, E., & Krathwohl, D. (Eds.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* New York: David McKay.

Chuska, K. (1995). Improving classroom questions: A teacher's guide to increasing student motivation, participation, and higher level thinking. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Ellis, K. (1993). Teacher questioning behavior and student learning: What research says to teachers. (Paper presented at the 1993 Convention of the Western States Communication Association, Albuquerque, New Mexico). (ERIC Document Reproduction No. 359 572).

Gall, M. (1970). The use of questions in teaching. *Review of Educational Research, 40,* 707-721.

Gall, M. (1984). Synthesis of research on teachers' questioning. *Educational Leadership, 42,* p. 40-47.

Leven, T. and Long, R. (1981). *Effective instruction.* Washington, DC: Association for Supervision and Curriculum Development.

Morgan, N., and Saxton, J. (1991). *Teaching, Questioning, and Learning.* New York: Routledge.

Rosenshine, B. (1971). *Teaching behaviors and student achievement.* London: National Foundation for Educational Research in England and Wales.

Sanders, N. M. (1966). *Classroom questions: What kinds?* New York: Harper & Row.

Stevens, R. (1912). *The question as a means of efficiency in instruction: A critical study of classroom practice.* New York: Teachers College, Columbia University.

Wilen, W. (1991). Questioning skills for teachers. What research says to the teacher. Third edition. Washington, DC: National Education Association. (ERIC Document Reproduction No. 332 983).

Wilen, W. and Clegg A. (1986). Effective questions and questioning: A research review. *Theory and research in social education, 14*(2), p.

153-61.

Descriptors: Academic Achievement; *Classroom Techniques; Discussion (Teaching Technique); Elementary Secondary Education; Learning; *Questioning Techniques; Teaching Methods; Wait Time

Home | Articles | Subscribe | Review | Policies

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

MacColl, Gail S. & White, Kathleen D. (1998). Communicating educational research data to general, nonresearcher audiences. *Practical Assessment, Research & Evaluation*, 6 (7). Available online: http://ericae.net/pare/getvn.asp?v=6&n=7. This paper has been viewed 8416 times since 11/13/99.

Communicating educational research data to general, nonresearcher audiences

Gail S. MacColl & Kathleen D. White
Program Evaluation and Methodology Division
U.S. Government Accounting Office

Parents, educators, school board members, and legislators all want to know "what works" and "what doesn't" in terms of educational programs and innovations. The reasons for their interest are obvious and worthwhile: first, they want to be sure that tax money is being spent on educational programs that provide a positive return in terms of student progress; second, they want to stay informed of trends in education so they know that their school districts are keeping up with the latest practices and programs.

This digest describes some of the problems in communicating with these audiences; it then provides helpful information on how researchers can best present data on educational practices that work and those that don't to these various audiences for maximum effectiveness, impact, and influence and to keep communication with these audiences open and valuable.

Problems in effective communication to general, nontechnical audiences

Accessibility

Most research on effective educational practices does not filter down to the people who contribute to or control funding. The main reason for this is that research reports on educational practices almost universally appear only in professional and academic journals or through other specialized sources.

The average reader wanting to learn about successful innovations in education is generally unable to locate such information, even after expending considerable effort. These kinds of reports are usually unavailable through popular periodicals or bookstore chains and rarely through more "serious" bookstores; in addition, they are not often found in or through local libraries, including those in large metropolitan areas.

Readability

In the rare event that a general reader gains access to materials about workable educational programs, three stylistic characteristics of these reports often make them unappealing: organization, terminology, and presentation of statistical data.

First, research studies are often *organized* in such a way as to hide major findings and conclusions in the text or present them only at the end. A related problem is that abstracts and introductions do not provide findings. Even diligent readers become discouraged by these factors because the usefulness of a report or study is not readily apparent.

Second, although the use of *technical terminology* often simplifies communication within a discipline, it creates an obstacle for policy makers, parents, and other interested readers, who usually are not trained in research or statistical techniques needed to understand an esoteric research study.

Third, many research studies use complex tables to summarize statistical data. These tables, like research jargon, are often difficult for interested parents and program funders to interpret, even with considerable effort.

Difficulties in reporting "what doesn't work"

Researchers often have difficulty in reporting on educational practices that don't work, generally for one or more of the following reasons:

- This question is often addressed only indirectly because most studies reports on something that works, only implying that its absence or the presence of its opposite doesn't work. Unless it is strongly and directly stated, the message that something is ineffective rarely comes out.
- No broad agreement exists about the meaning of "a practice doesn't work." First, "practice" is defined either very broadly or in ideal terms, not in any generally accepted way; second "doesn't work" could have any of several meanings:
- The "practice" is difficult or impossible to implement as intended.
- It has not succeeded in most places where it has been tried.
- It is associated with generally negative results or minimally positive ones.
- It generates fewer positive results than alternative practices.

- Evidence that a practice doesn't work is rarely unequivocal, in part because the results it generates may change in different contexts or with slight modification.

Technical weaknesses that limit usefulness

Assuming that the other problems are overcome, several weaknesses can occur in the research itself to limit its value to those funding, evaluating, or deciding on the use of new educational programs:

- *Design constraints*--Many reports are based on single case studies. This factor limits a study's applicability

and its generalizability.

- *Insufficient demographic data and contextual background*--When this weakness appears in an article, general readers becomes concerned about the applicability of a practice in their situations.
- *Lack of findings in terms of student progress*--The few studies that include these data used limited outcome measures, such as norm-referenced tests and SAT scores to measure classroom achievement.
- *Little guidance in improving practices in meta-analyses*--Although often effective in summarizing a number of studies, meta-analyses in these studies generally omit important details for determining the value of individual studies.
- *Policy statements disguised as objective research*--Readers must exercise care to separate useful studies from these, which are supported by both shaky assumptions and selective data; unfortunately, with the growing popularity of the Internet, these will probably be found with increasing frequency on the Internet.

How to increase the value of research studies to a wider audience

Most of the recommendations for making research data more useful to more people are simple, relatively easy to accomplish, and based on common sense.

In general, the primary things to focus on are the needs of the audience. Researchers must remember that, in order for their data to be most useful, they have to be accessible and understandable to people with vested interests in the education process: parents, teachers, legislators, school board members. These audiences either pay for, deliver, or fund education programs, and each wants the best ones available.

Researchers uncovering and reporting on programs and practices that work need to distribute their findings as widely, clearly, and efficiently as they can; otherwise their efforts do not create the levels of benefits for the discipline of education that they might.

More specifically, when reporting study results to nontechnical audiences, researchers should keep the following suggestions in mind:

- Summarize the findings in plain language at the beginning of the report. Most nonresearchers appreciate getting the meat of the matter quickly without having first to trim away the fat.
- Present the information in a manner that allows it to be absorbed quickly. As with most of us, even the most interested general readers have time constraints. The more a researcher can do to help readers overcome this problem, the more that he or she will benefit the future of education.
- Provide more detailed material later in a report for those wanting it, but not in place of the summary data.
- Communicate through channels that reach the general public.

To accomplish these goals, researchers will have to learn how to creatively present their findings not only to reach more general readers but to appeal to them too. This requires several steps:

- Simplifying language so that readers without backgrounds in research or statistics can readily understand the content of a report.
- Creating simple tabular material that readers can more easily interpret than dense statistical tables sometimes found in scholarly research journals.
- Incorporating inviting graphics into materials intended for general audiences. These tend to encourage reading and help reader understanding of the material.
- Enlisting the aid of journalists and other communicators who can help both in designing the information for mass consumption and in placing the information in media that the general reader will see.
- Publishing on the Internet, an extraordinarily powerful tool for making information accessible to a wide audience.

- Making certain that the research supports your conclusions, that the work contributes to advancing the level of education, and that a critical eye was used to examine the purpose, the objectivity, and the methodology behind the study.

References

Good, R. G. & Wandersee, J. H. (1991). "No Royal Road: More on Improving the Quality of Published Educational Research." *Educational Researcher,* 20 (8), 24-25.

Newmann, F. M. & Wehlage, G. G. (1995). *Successful School Restructuring: A Report to the Public and Educators by the Center on Organization and Restructuring of Schools.* Madison, WI: Center on Organization and Restructuring of Schools.

U. S. General Accounting Office. (1995). *Program Evaluation: Improving the Flow of Information to the Congress* (GAO/PEMD-95-1). Washington, DC.

U. S. General Accounting Office. (1996). *Schools and Workplaces: An Overview of Successful and Unsuccessful Practices* (GAO/PEMD-95-28). Washington, DC.

Descriptors: *Communication (Thought Transfer); *Educational Research; Higher Education; *Readability; *Research Reports; *Researchers; Technical Writing; *Writing for Publication

Home | Articles | Subscribe | Review | Policies

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

Search: title Go

Calkins, Lucy, Kate Montgomery & Donna Santman (1999). Helping children master the tricks and avoid the traps of standardized tests. *Practical Assessment, Research & Evaluation*, 6(8). Available online: http://ericae.net/pare/getvn.asp?v=6&n=8. This paper has been viewed 13244 times since 11/13/99.

Helping Children Master the Tricks and Avoid the Traps of Standardized Tests

Lucy Calkins, Kate Montgomery, and Donna Santman

Adapted from A Teacher's Guide to Standardized Reading Tests. Knowledge is Power (1998) by Lucy Calkins, Kate Montgomery, and Donna Santman, Portsmouth, New Hampshire: Heinemann.

Introduction

Children can improve and change their test-taking habits if they are taught about their misleading work patterns. Teaching children about the traps they tend to fall into may well be the most powerful, specific preparation teachers can give them for the day of the test. By studying the habits of young test takers, we uncovered some of their common mistakes. This Digest lists some of these mistakes and suggests several teaching strategies that may be useful to teachers who are preparing their class to take standardized tests.

Use the Text to Pick Your Answer

When it comes to choosing an answer, many children are much more likely to turn to their own memories or experiences than to the hard-to-understand text for their answers. This issue becomes even more difficult when the passage is an excerpt from a text with which the students are familiar. Many new reading tests use passages from well-known children's literature, including those stories that have been made into movies. In this case, many students justify their answers by referring to these movies or their memory of hearing the story when they were younger.

While these personal connections are helpful if the student is at a complete loss for an answer, it's essential for

children to understand that relying on opinions, memories, or personal experience is not a reliable strategy for finding answers that a test maker has decided are correct. Clearly, many questions asked on the tests require prior knowledge to answer, but the problem comes when students rely exclusively on that prior knowledge and ignore the information presented in the passage. Some things that teachers may wish to do in order to help their students avoid making this mistake include the following:

- Teach students to underline parts of the passage that might be asked in the questions
- Help children develop scavenger-hunt-type lists of things to look for as they read the passages by having them read the questions first
- Teach students to find out how many questions they can hold in their minds as they read the passage
- Show children how to fill in all the answers on each test booklet page before filling in the corresponding bubbles on the answer sheet
- Teach children ways to mark the passage in order to make it easier to go back to find or check specific parts - these include writing key words in the margins and circling or underlining
- Show students how to use an index card to block out distracting print or to act as a placeholder
- Retype familiar or easy text to look as daunting and dense as the test passages to give children confidence and experience in the test format.

Sometimes It's Helpful to Refer to Your Own Life Experiences

In the reading comprehension sections of a reading test, children must find evidence in the passages to support their answers. Yet, there are parts of many reading tests where the only things students can rely on are their own previous experiences. In these sections, students are asked to choose the correct spelling of the underlined word or to choose the word whose meaning is closest to that of the underlined word.

Often students prepare for these sections of the tests by taking practice tests and then going over the answers. However, it is highly unlikely that any of the same words would appear on the actual test. Therefore, teachers may wish to impress upon children the importance of creating a context for the variety of words that may be found on the test by relating those words to their own personal reading experiences. In order to facilitate that thinking process, teachers may wish to help children ask themselves such questions as "Have I seen this word before in a book?" "Where have I heard that before?" or "What words or events usually happen around this word?" while they are answering vocabulary or spelling questions.

Learn to Read the Question

It is always assumed that if children have reading troubles, their wrong answers stem from difficulty reading the passages. However, this is not always the case. Sometimes, reading the questions, a much less familiar task, can prove to be the greatest reading challenge for the students. This is because questions such as "How was the central problem resolved?" or "Which statement is NOT true about the narrator?", are not the types of questions children are asking themselves and each other about the books they read.

Studying various types of questions can be a helpful practice to future test takers. This can be done by searching through practice tests and making lists of the types of questions. Although the questions will be different on the day of the test, this exercise may familiarize students with the types of questions that are asked on standardized tests.

Choose the Answer to the Question

Sometimes children choose their answer by finding the first answer choice that matches something in the text. Unfortunately, by not considering what the question was actually asking, they are tricked into choosing the wrong answer simply because it may state a fact that was included in the story.

One teaching strategy that can help students avoid this mistake is to present a text with questions in a standardized test format. With a partner, the child should figure out what the different questions are asking, and write down their paraphrased versions. Many times children will be surprised at how different their paraphrasing is from what the question is actually asking. It may be a good practice for teachers to look at the different paraphrasings with the class and discuss which interpretations would help the members of the class and which would lead them astray. This allows students to strengthen their skills at finding the true meaning of the questions.

Risk an Unfamiliar Choice

Frequently, students avoid choosing an answer simply because it contains an unknown word even when they know the other choices are probably wrong. Thus, teachers should advise students not to overlook the possibility that the answer which contains the unfamiliar word may be the correct choice. Teachers often try to teach children a way of narrowing down the answer choices through a process of elimination. Despite the fact that this process can be very helpful, many students eliminate two possibilities and then, from the last two, just sort of pick one. They don't, it seems, try to figure out a reason to choose one over the other. They seem to wrongly assume that the two choices left are equally possible. However, teachers should teach students that thoughtful elimination between the two last possibilities can lead to the correct choice.

Check Your Answers

After the harrowing ordeal of taking a standardized test, the last thing that students usually want to hear coming from their teacher is "Did you check your answers?" Frequently, the biggest reason kids hate checking answers is because they have only one strategy for doing so: opening their test booklets to the first passage and beginning again. To them, checking answers means taking the test again. However, that does not have to be the case. There are a variety of different strategies that students can use for selectively going back through the test and reconsidering answers. One of these strategies is teaching children to only check the problems of which they were unsure. It is unnecessary to return to questions about which students feel fairly confident. Students can keep track of the troublesome questions while they are actually taking the test. They can do this in several different ways: jotting down the numbers of the questions on a separate sheet of paper, circling the numbers in the test booklet, etc. Students should also know that it is okay to take a short break (stretching in their seats, bathroom/drink break) before going back and checking the answers. This will give them a chance to clear their minds a little bit. Most importantly, students should be taught to attempt to check the answers to the troublesome questions using a new strategy so that they may avoid reusing possibly faulty problem-solving methods.

Setting the Tone for Test Day

Although teachers may do their best to prepare their students for standardized tests, every teacher has stories of children dissolving into tears on the day of tests. Even if their feelings aren't so obvious, all children feel the pressure of doing well. Be sure you don't add to the pressure by over reacting to small deeds of misbehavior or by over emphasizing the fact that today is a testing day.

Suggested Readings

Calkins, L., Montgomery, K. and Santman, D. (1998). *A Teacher's Guide to Standardized Tests. Knowledge Is Power.* Portsmouth, NH: Heinemann.

Mitchell, R. (1992). *Testing for learning: How new approaches to evaluation can improve American schools.* New York: The Free Press.

Perrone, V. (Ed.). (1991). *Expanding student assessment.* Alexandria, VA: ASCD.

Shepard, L. (February, 1995). Using assessment to improve learning. *Educational Leadership*, 38-43.

Descriptors: Standardized Tests

Home | Articles | Subscribe | Review | Policies

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

Matter, M. Kevin (1999). Strategies for improving the process of educational assessment. *Practical Assessment, Research & Evaluation*, 6(9). Available online: http://ericae.net/pare/getvn.asp?v=6&n=9. This paper has been viewed 21336 times since 11/13/99.

Strategies for Improving the Process of Educational Assessment

M. Kevin Matter
Cherry Creek (CO) Public Schools

Test administration is an essential part of the educational assessment process, yet it often does not receive enough attention. Because teachers and principals are concerned with many components of the testing process, it is important for the assessment office to focus attention on test administration. This digest presents seven strategies that the assessment director may employ to improve test administration practices. These strategies highlight clear communication, the responsibility of the Building Test Coordinator, and rewarding and reinforcing quality. The administration process from school staff's perspective and the needs of the assessment office are both addressed.

Communicate--Communicate--Communicate

Parents and teachers rarely learn how results are used to improve curriculum, instruction, or individual student learning plans. Assessment offices and school districts have a responsibility to provide them with that information. Develop a year-long communication plan for school staffs, parents, and the community. It is important for everyone affected by the assessment process to be continually informed. They should know what tests are being administered, the purpose of the tests, what the past results show, and how the current results are used to improve student performance.

Tailor the information to fit the needs of the audience. Providing teachers and principals with test administration checklists, manuals, and reports to meet the assessment office needs for standardization and efficiency is not enough. They should be provided with information that meets their needs as customers of the test: how will the test impact their students, curriculum, and district? Briefly communicating to them the assessment impact reinforces the teamwork that is needed to ensure an assessment system that is both used and useful.

Designate a Building Test Coordinator

The key to both administration and processing quality is a team which includes a Building Test Coordinator at each school. This person is responsible for administering test materials, overseeing the test administration process, and providing the assessment office with quality materials. The Building Test Coordinator works as a liaison between the assessment office and the school.

Appoint qualified staff to assist the Building Test Coordinator with materials and with administration and scoring issues. This additional help will free the Building Test Coordinator to maintain ongoing communication with the assessment office much more easily. Have someone with a more flexible schedule assume responsibility for issuing materials. Possible choices for this position are the clerical staff, a teacher assistant, or a counselor. The administrative/scoring role may be filled by either a teacher or an administrator, as long as the person is knowledgeable about or will be trained in the technical and instructional issues of assessment.

Do not use the principal as Building Test Coordinator. Although it is important that the principal remain informed and involved in the assessment process--especially regarding deadlines and requirements -- the best role for the principal is to support the Building Test Coordinator by providing extra help and resources.

Meet with ALL Building Test Coordinators

Require all Building Test Coordinators to attend a brief overview meeting with the assessment staff. To keep the Building Test Coordinator informed, regularly share what works in the school or district, such as providing extra clerical time before and after testing days.

Do not send test materials through the mail. Provide all test materials at the meeting (except test booklets). Walk through all expectations (coordinator and teacher checklists, materials list, materials check-out sheet, administration directions) at the meeting.

Make the Building Test Coordinator personally responsible for the test materials. Before testing begins, communicate that the Building Test Coordinator must ensure that the materials provided meet the acceptable standards. Require the Building Test Coordinator to personally deliver the answer sheets after the testing (or arrange area "drop-off" locations around your district). Schedule a time for check in.

Develop a process to inspect the test result materials. Whenever possible, provide the Building Test Coordinator with options for the school (e.g., hiring part-time staff to prepare the completed materials). Explain that unacceptable materials will either be returned to the school, or schools will be charged for processing time. Use area check-in locations throughout the district, as needed.

Stress quality of test result materials. Explain the consequences of poor quality of materials returned by the Building Test Coordinator. Also emphasize the consequence of a particularly long turnaround time. Provide examples of what "good materials" look like (answer sheets completed correctly, header sheets completed, etc.). Explain that good input at the teacher/school level can alleviate hours of time at the assessment staff level.

Design Processes to Reward Quality

Recognize a job well done. Find out what is rewarding to the Building Test Coordinator and do that! Examples of inexpensive tokens of appreciation include:

- Gift certificates from a book store
- Certificates of appreciation

- Letters of thanks to supervisors
- Thank You party
- Feedback on materials returned

Delaying the reports or results from one testing location because of problems with other teachers or schools "punishes" high quality work. If feedback deadlines are observed, schools will be quickly rewarded for their efforts.

Use "Quality" Techniques

Use an effective system of deadline dates. Good procedures include several "waves" of processing and reporting, with deadline dates determined by the time that is needed to properly collect and prepare materials for delivery to the testing office. Test coordinators will know the deadlines and understand the relationship between the date and quality of how materials were submitted for processing and the date the results are received.

Remember the Golden Rule. Assessment offices may be viewed by teachers and principals as "the enemy" if practices involve high stakes accountability and unfair treatment. To counteract these perceptions, assessment processes must be developed that involve the "user/customer" throughout the entire process, not just at the end.

The assessment office must design goals, processes, and procedures with the following in mind:

Information: All information that is provided must be timely and understandable. Materials should meet the needs and expectations of the user.

Responsiveness: Assessment staff must be accessible at times that are conducive to the culture of the school and the time demands placed on the teacher, the principals, and other staff.

Input: Ask for feedback whenever possible, particularly when the user is qualified to comment on the quality of the material. Act on this feedback, making the necessary improvements.

Teamwork: Teachers, principals, parents, and assessment offices must work together. Communicate the idea that performance at one school affects other schools in the district.

Rapid turnaround: Reward schools by providing rapid processing and reporting of results. Late is almost identical to never with assessment results.

Reports: Spend the additional time and resources necessary to customize reports for each audience. The payoff for reports that are understandable is actual use of the results.

Useful and usable information: Create staff development and training for teachers, principals, school staff, and parents that is focused on assessment results they need and value.

High standards: Demonstrate that the high standards that apply to others apply to the assessment office processes and procedures as well. Take prompt action to rectify identified problems.

Continual Improvement in Processes

Efforts to improve administration, processing, and reporting take several years. Plan for incremental steps to change behavior by rewarding and reinforcing quality results. Keep a log of good practice ideas; use this to reduce variation and problems when using a particular process. Be positive, but expect new problems to occur even as others are reduced.

Involve the entire assessment staff in the planning process, as well as key representatives from the various internal and external audiences. Allow assessment offices to be seen as "a part of," rather than "apart from" the schools and teachers. Consider the point of view of all involved in the assessment process.

Communicate continuously with assessment staff, building administrators, central office, and as much as possible with the test coordinators and teachers. Convince them of the benefits of improvements in the entire assessment process--more usable information, at a lower cost to the taxpayer.

Additional Reading

Arkley, H., and others. (1988). *Assessing student performance for school improvement*. Springfield, IL: Illinois State Board of Education. ED 300 418.

Caswell, M. S., & Roeber, E. D. (1982). *Reporting test results to the school board. Using and reporting test results, monograph #5. Steps in the right direction*. Lansing, MI: Michigan Educational Assessment Program (MEAP). ED 246 119.

Cuban, L. (1984). Transforming the frog into the prince: effective schools research, policy, and practice at the district level. *Harvard Educational Review, 54(2)*, 129-151.

Lazarus, M. (1982). *Evaluating educational assessment programs*. Arlington, VA: American Association of School Administrators. ED 226 414.

Nichols, J. O. (1990). The role of institutional research in implementing institutional effectiveness or outcomes assessment. *Association for Institutional Research, 37*, 7. ED 323 849.

Descriptors: not implemented

Home | Articles | Subscribe | Review | Policies

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

Search: title Go

Quellmalz, Edys, Patricia Schank & Thomas Hinojosa & Christine Padilla (1999). Performance assessment links in science. *Practical Assessment, Research & Evaluation*, 6(10). Available online: http://ericae.net/pare/getvn.asp?v=6&n=10.

Performance Assessment Links in Science (PALS): http://pals.sri.com

Edys Quellmalz, Patricia Schank, Thomas Hinojosa, and Christine Padilla
SRI International, Center for Technology in Learning

Understanding the central role that performance assessment plays in standards-based reform, educators are seeking ways to use these assessments to test student learning. Education agencies need pools of performance tasks to use in their student assessment programs and in evaluations of state and federally funded programs. Reform projects need standards-based assessment, too, as do teachers who are trying to implement reforms. Experience indicates, however, that the level of effort and costs of developing performance assessment tasks are very high (Quellmalz, 1984).

To meet the need for innovative approaches for sharing exemplary assessment resources, to facilitate the development of new ones, and to further understand how the use of standards-based performance assessment can advance educational reform, SRI International is developing Performance Assessment Links in Science (PALS), an on-line, standards-based, interactive resource bank of science performance assessments. Coupled with the development of the resource bank is a program of research on effective use of these resources.

This digest describes work-in-progress. SRI is currently seeking organizations to participate in implementation studies involving the use of PALS. The "ideal" professional development model would be for an organization or group of schools to want to focus on classroom applications of science performance assessment and to use the PALS resources in an initial professional development institute, followed by several school-year meetings where teachers would bring samples of student work produced in response to PALS investigations, with subsequent work on designing additional classroom science assessments. For the accountability studies, SRI will work with a few assessment programs from states, districts, or specially funded programs that are interested in using the on-line,

secure tasks in the Accountability Pool. Assessment programs might use PALS to search, select, and plan assessment administrations and/or conduct on-line rater training and scoring sessions.

Performance Assessment in School Reform

Reform-minded programs have set out to develop alternative forms of student assessment that call for students to construct rather than select responses. Performance assessments are generally valued for testing students' deep understanding of concepts and inquiry strategies, for making students' thinking visible, and for measuring skills in communicating about their science knowledge. In addition, performance assessments can present authentic, real-world problems that help students to show they can apply academic knowledge to practical situations. On the other hand, performance assessments are time-consuming and costly to develop, logistically demanding, and of questionable utility if not developed and scored according to sound measurement methods.

Technology in Assessment Reform

Technology offers a powerful strategy for increasing the ease with which educators can access and use standards-based student assessment. Technology can be used to efficiently archive numerous assessments for ready browsing. Currently, some sets of performance assessment tasks are distributed on CD-ROM, and there are plans to place released test items on the Internet, although these resources are not yet coordinated or easily accessible to programs and schools. Electronic networks can go beyond storage by supporting the growth of a community of colleagues and leveraging expertise, regardless of geographic location and institutional base. Networks can offer templates and guidelines and support collaborative development and on-line conversations about the alignment of tasks with standards, quality of tasks and student work, on-line training, scoring, and standards-setting sessions. Technology can also support simulated investigations and data collection and analysis of student responses. The exponential advances taking place in technology promise to revolutionize assessment practices and education reform (Kozma & Schank, 1998; Quellmalz, 1999).

Technology in Professional Development

New professional development models are needed to provide teachers with greater opportunity to access, discuss, incorporate, and co-construct assessment resources and other reform-based materials (Little, 1993). Current efforts find it difficult to reach many teachers and to maintain discourse, and teachers have difficulty implementing new ideas back at their schools. Technology can help provide mechanisms for teachers to overcome their isolation and make more effective use of their time spent on professional growth.

Performance Assessment Links in Science (PALS)

PALS provides an on-line assessment resource library designed to serve two purposes and user groups: (1) the accountability requirements of state education agencies and specially funded programs, and (2) the professional development needs of classroom teachers. SRI is developing PALS under a grant from the Instructional Materials Development (IMD) program within the National Science Foundation with two primary goals:

(1) To develop a two-tiered on-line performance assessment resource library composed of performance assessment tasks for elementary, middle, and secondary levels. One tier will be a password-protected, secure Accountability Pool of science performance assessments for use by state assessment programs and systemic reforms. The second tier is for use by teachers and professional development organizations. The Professional Development Pool provides performance-based science assessments that have been used successfully in large-scale (state or national) assessment programs and have been released.

(2) To evaluate the effectiveness of policies, implementation models, and technical quality requirements for the use

of the two tiers of PALS.

In our design for PALS, experienced assessment programs can contribute standards-based science assessment tasks with documented technical quality to the PALS on-line library. The Accountability Pool will be composed of password-protected, secure tasks accessible only to approved assessment program staff. Assessment programs can thus share their resources and enjoy a much larger pool of science performance assessments to use or adapt for their testing administrations. The PALS resource library can provide large, continually updated collections that can support efficient searching, selection, and printing.

The Professional Development Pool contains resources that are of documented technical quality and have been released for access by teachers and professional development groups. Pre-service and in-service programs, for example, can reach teachers in geographically distributed and remote locales, resulting in great savings of travel and materials expenses. On-line guidelines and templates can support classroom use of science performance assessments. Teachers may administer the performance tasks as part of their classroom assessments, adapt them, or use them as models for developing similar investigations.

To help users design test administration forms that cover important science standards, the on-line system provides assessment planning charts (Stiggins, 1994). PALS automatically produces an assessment planning chart to display coverage of standards by the performance tasks selected by the user.

The science performance assessment library includes the scoring rubrics designed to judge the quality of student responses to a task or set of tasks. To bring meaning to the scoring rubrics, a library of exemplars of scored student work is included.

Rater training materials are not routinely published by assessment programs. SRI has developed specifications for on-line rater training and scoring so that each agency wishing to take advantage of the PALS system can convert traditional, stand-up training procedures and calibration to written form, assemble training packets, and test their effectiveness in on-line delivery.

An essential component of PALS is the documentation of the technical quality of each task in the library. PALS contains the technical quality indicators provided for the field-tested science tasks. Since the resource library will be stocked only with tasks that have survived a systematic development process, the tasks and rubrics will also have been subjected to content validity and bias reviews.

Issues and Expansion of the System

The project will be addressing a number of issues. One will be the procedures for identifying the science standards that tasks are designed to test and the nature of the groups that will make the alignment judgments. Another issue relates to policies for accessing the resources that have been developed with funds from different sources and that represent assessment materials being distributed by various organizations. A third issue relates to the criteria and procedures for reviewing tasks submitted for inclusion in the bank. Finally, the system must develop strategies for operating, maintaining, and expanding the resources.

PALS uses technology to efficiently archive numerous assessments for ready browsing by teachers, and goes beyond mere storage of assessments to support cross-links with standards, tasks, scoring criteria, and annotated student work. These resources can be shared by assessment programs, allowing them to have access to a large pool of performance assessments to use or to adapt for their testing administrations. Teachers can administer the tasks as part of their classroom assessments, adapt them or use them as models for developing their own investigations, and contribute their adaptations to the resource bank for other teachers to use.

As an on-line library alone, PALS is a valuable set of resources. However, the growing body of professional development research and the results of our pilot study suggest many benefits of integrating more collaboration support into PALS. SRI plans to integrate an online meeting/discussion component to help the community members leverage expertise, regardless of geographic location. This component could be used to support collaborative development and on-line conversations about the alignment of tasks with standards, the quality of the assessment tasks and student work, rater training and scoring, and standards setting. The authors believe that, by taking advantage of new models of professional development that include innovative digital technologies, PALS will provide excellent professional development opportunities for teachers.

References

Kozma, R., & Schank, P. (1998). Connecting with the twenty-first century: Technology in support of education reform. In D. Palumbo and C. Dede (Eds.), *Association for Supervision and Curriculum Development 1998 yearbook: Learning with technology.* Alexandria, Virginia: Association for Supervision and Curriculum Development.

Lieberman, A. (1996, November). Creating intentional learning communities. *Educational Leadership*, 51-55.

Little, J. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis*, 15(2), 129-151.

Quellmalz, E. S. (1984). Designing writing assessments: Balancing fairness, utility, and cost. *Educational Evaluation and Policy Analysis*, 6, 63-72.

Quellmalz, Edys (1999). The role of technology in advancing performance standards in science and mathematics learning. In Kathleen Comfort (Ed.), *Advancing standards for science and mathematics learning: Views from the field.* Washington, DC: AAAS.

Stiggins, R. J. (1994). *Student-centered classroom assessment.* New York: Macmillan College Publishing Company.

Descriptors: *Information Dissemination; *Online Systems; *Performance Based Assessment; Professional Development; *Resources; *Sciences; *Standards

Home | Articles | Subscribe | Review | Policies

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal. ISSN 1531-7714

Scriven, Michael (1999). The nature of evaluation part i: relation to psychology. *Practical Assessment, Research & Evaluation*, 6(11). Available online: http://ericae.net/pare/getvn.asp?v=6&n=11. This paper has been viewed 17661 times since 11/13/99.

The Nature of Evaluation Part I: Relation to psychology

Michael Scriven,
Claremont Graduate University

The discipline of evaluation is devoted to *the systematic determination of merit, worth, or significance*. It is divided into fields according to the type of entity evaluated--for example, program evaluation, or personnel evaluation--and there are more than twenty of these recognized fields of evaluation. Some specific aspects of evaluation methodology have been developed to solve problems of evaluation in only one or a few of these fields (e.g., bias control in panel selection, systematic side-effect identification in program evaluation, road-testing techniques in product evaluation). However, the underlying logic of the process of evaluation--for example, the difference between merit and worth, or between grading and ranking--and a substantial portion of its general methodology (e.g., techniques of measurement, causality determination, applying the requirement of informed consent) are shared across all or many of these fields. Many of these general techniques (the 'general methodology') come from the applied social sciences, and are learnt by students in the normal course of education in those fields. But the logic of evaluation has been developed for and applies only to evaluation; and the field-specific methodologies of evaluation must also be mastered in order to deal with evaluation in the fields to which they apply. Teaching evaluation therefore focuses on these evaluation-specific topics, the general logic and the special methodologies.

This article addresses the role of evaluation, the basic logic, and a description of how the field is structured. A separate article describes some of the basic *logic of evaluation* skills and methodological skills that need to be mastered..

Evaluation in applied psychology

Just as it was previously found that a good grasp of probability and statistics had become an essential tool for a great deal of work in applied psychology, so we now find that a knowledge of the logic of evaluation and of some of its

Home | Articles | Subscribe | Review | Policies

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

Search: title Go

Scriven, Michael (1999). The nature of evaluation part ii: training. *Practical Assessment, Research & Evaluation*, 6(12). Available online: http://ericae.net/pare/getvn.asp?v=6&n=12. This paper has been viewed 17574 times since 11/13/99.

The Nature of Evaluation Part II: Training

Michael Scriven,
Claremont Graduate University

An earlier article addressed the role of evaluation, the basic logic, and a description of how the field is structured (Scriven, 1999). This article describes some of the basic *logic-of- evaluation* skills and some of the basic methodological skills that need to be mastered in order to practice the art and science of evaluation.

Much work in the Big Six evaluation fields - program, personnel, performance, policy, proposal, and product evaluation - falls within the area of applied social psychology, and much of that -- e.g., the evaluation of large social interventions -- would be impossible without training in the methods and mathematics that foundations requirements in graduate psychology now cover. But there is at least one other completely different kind of reason for thinking the connection between psychology and evaluation is an intimate one, namely the highly specific phenomena of reactions to evaluation by those being evaluated and those for whom the evaluation is done. Dealing with these is an important part of developing applied skills in evaluation. However, the standard training provided in standard psychology programs will not put the graduate in a position where s/he can deal competently with common phenomena in evaluation. Nor should this be regarded as a matter for clinical training, although it is related, and although there are times when the phenomenology comes very close to the clinically relevant level.

Logic-of-evaluation skills

The following list indicates some of the topics from the logic of evaluation that must also be dealt with in some detail.

1. Understanding the differences and connections between evaluation and other kinds of research and investigation, especially: description, classification/diagnosis, generalization, prediction, explanation, justification, and recommendation. Hence, understanding the different types of research design and data inputs required for each of

these.

2. Understanding the difference between:

i. grading, ranking, scoring, and apportioning (the basic evaluative procedures);
ii. merit (or quality), worth (or value), and significance (or importance)--the basic evaluative predicates.

Hence, understanding the differences between investigative designs aimed at establishing conclusions of these (theoretically 12, but actually about 6) different types. Specific case: understanding the function of 'significance levels' in statistics by contrast with significance determination in scientific or social research.

3. Understanding the arguments that purported to establish the impossibility of scientific demonstrations of evaluative conclusions, and the reasons they failed. (The 'Science is only descriptive' argument; the 'Values are always subjective' argument; the 'Naturalistic fallacy' argument.) Understanding why the usual arguments against value-free science also fail (the 'Scientists show their values in choosing their field/research problems' argument; the 'Science is used for good or bad purposes' argument.) Understanding why these arguments are not just philosophical exercises but reflections of common client/audience confusions that need to be dealt with.

4. Understanding the difference between (I) holistic (black box) evaluation (ii) analytic evaluation; and between the three kinds of analytic evaluation--dimensional, component, and theory-driven evaluation; and how to choose between them in approaching a particular evaluation problem.

5. Understanding the formative/summative distinction, and some of the arguments for thinking that a third category should be included to make up a complete classification of all evaluations.

6. Understanding the nature of needs assessment and its difference from market research; and how to design a valid needs assessment.

7. Understanding the logic of checklists, especially the difference between checklists of (I) desiderata and (ii) necessitata; and the logical requirements for validity of each kind.

8. Understanding the differences and connections between objectivity and: (I) bias, (ii) preference/valuing/valencing; (iii) commitment; (iv) expertise. The fallacy of irrelevant expertise in selecting evaluators. The views of realists and constructivists about objectivity.

9. Understanding the range of evaluation approaches on the scale from fully distanced to highly interactive, and the 'off-scale' entries of description and evaluation training; all with their attendant advantages and disadvantages.

10. Understanding the difference between the kind of evidence required to establish causation and that required to demonstrate culpability.

11. Understanding how and why evaluation developed from (I) a practice to (ii) a highly skilled/professional practice to (iii) a field-specific discipline and finally (iv) to a transdiscipline.

12. Understanding how evaluation theory developed from the primitive identification of evaluation with monitoring to its present complex form, including goal-free evaluation; and understanding some of the leading positions taken by influential theorists along the way and today.

Methodological Skills

The following is a list of a list of some methodological skills of great importance in evaluation which are rarely, if ever, covered in the core curriculum of psychology graduate curricula.

1. The Key Evaluation Checklist approach, including details of how to determine the five mainline checkpoints (Outcomes, Process, Costs, Comparisons, Generalizability).

2. Meta-evaluation procedures; the four approaches (recheck, redo, do differently, special checklists).

3. Cost analysis, especially of non-money costs.

4. Skills from qualitative research, notably the determination of causality in non-experimental research, e.g., in medicine (the lung cancer case and the paresis case), and in history (the causes of unpreparedness at Pearl Harbor).

5. Some intradisciplinary skills, especially theory evaluation.

6. How to identify relevant values for a particular evaluation and deal with highly controversial values and issues e.g., in evaluating family planning programs, or in dismissal procedures.

7. How to report to non-peer clients, stakeholders and audiences, especially using non-text media.

8. The psychology of evaluation, especially managing evaluation anxiety.

9. Some field-specific skills, in e.g., technology assessment, personnel evaluation, business evaluation, non-profit management, developmental evaluation, proposal evaluation, evaluative questionnaire design, etc.

Additional Reading

Chelimsky, E and Shadish W.R (eds) (1997) *Evaluation for the 21st Century : A Handbook.* Sage Publications. [amazon]

Joint Committee on Standards for Educational Evaluation (1998). *Program Evaluation Standards : How to Assess Evaluations of Educational Programs*. Corwin Press. [amazon]

Scriven, M. (1991) *Evaluation Thesaurus* 4th edition. Sage Publications. [amazon]

Scriven, Michael (1999). The Nature of Evaluation Part I: Relation to psychology. Practical Assessment, Research & Evaluation, 6(11). [Available online: http://ericae.net/pare/getvn.asp?v=6&n=11].

Shadish W. R. (Chair) (1998) *Guiding Principles for Evaluators*. A Report from the Americian Evaluation Association Task Force on Guiding Principles for Evaluators. [available online http://www.eval.org/EvaluationDocuments/aeaprin6.html].

Shadish, William (1998). Some Evaluation Questions. *Practical Assessment, Research & Evaluation*, 6(3). [Available online: http://ericae.net/pare/getvn.asp?v=6&n=3].

Descriptors: *Evaluation Methods; Evaluation Problems; Evaluation Utilization; *Evaluators; Program Evaluation; Research Methodology; Theories; *Training

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

Rudner, Lawrence M. & William D. Schafer (1999). How to write a scholarly research report. *Practical Assessment, Research & Evaluation*, 6(13). Available online: http://ericae.net/pare/getvn.asp?v=6&n=13. This paper has been viewed 32496 times since 11/13/99.

How to Write a Scholarly Research Report

Lawrence M. Rudner, ERIC & University of Maryland
William D. Schafer, University of Maryland

Researchers communicate their results and help accumulate knowledge through conference papers, reports, on-line journals and print journals. While there are many rewards for having research disseminated in a scholarly outlet, the preparation of a good research report is not a trivial task.

This article discusses the common sections of a research report along with frequently made mistakes. While the emphasis here is on reports prepared for scholarly, peer-reviewed publication, these points are applicable to other forms of research reports. Dissertations and theses, for example, provide more detail than scholarly publications yet they adhere to the same basic scientific writing principles. Since all scientific research involves observation, description and analysis, points made in this article are applicable to historical and descriptive, as well as to experimental, research.

More detail can be found in the *Publication Manual of the American Psychological Association* (APA, 1994), proposed revisions to the manual (Wilkinson and Task Force on Statistical Inference, 1999), and many research methods textbooks (cf. Gay and Airasian, 1999). For general suggestions on publishing research, see Thompson (1995) and some of the articles and books also cited therein.

FIRST STEPS IN WRITING A RESEARCH REPORT

You should constantly think about writing your report at every stage of your research activities. The sections of the research report discussed next in this article come from the most-cited style source for educational and psychological literature - the *Publication Manual of the American Psychological Association* (APA, 1994). The *Publication Manual* provides detailed information about the entire process of publication -- from organizing,

writing, keying and submitting your manuscript, to seeing the accepted manuscript through production and publication. Of special interest in the fourth edition are the updated sections on reporting statistics, writing without bias, preparing manuscripts with a word processor for electronic production and publishing research in accordance with ethical principles of scientific publishing. You should have a copy.

Plan your report to focus on a single important finding or highly related group of findings. In the process of analyzing your data, you probably uncovered many relationships and gained numerous insights into the problem. Your journal article submission, however, should contain only one key point. The point should be so fundamental that you should be able express it in one sentence or, at most, in a paragraph. If you have several key points, consider writing multiple manuscripts.

When writing your manuscript, keep in mind that the purpose is to inform the readers of what you investigated, why and how you conducted your investigation, the results and your conclusions. As the investigator and writer, your job is simply to report, not to convince and usually not to advocate. You must provide enough detail so readers can reach their own conclusions about the quality of your research and the veracity of your conclusions.

SECTIONS OF YOUR REPORT

Title - It is important that the title be both brief and descriptive of your research. Search engines will use the title to help locate your article. Readers make quick decisions as to whether they are going to invest the time to read your article largely based on the title. Thus, the title should not contain jargon or vernacular. Rather, the title should be short (generally 15 words or less) and clearly indicate what the study is about. If in doubt, try to specify the cause and effect relationship in your key point. Avoid trite and wasteful phrases such as "A study of ..." or "An investigation to determine ..."

Abstract - The abstract serves two major purposes: it helps a person decide whether to read the paper, and it provides the reader with a framework for understanding the paper if they decide to read it. Thus, your abstract should describe the most important aspects of the study within the word-limit provided by the journal. As appropriate for your research, try to include a statement of the problem, the people you studied, the dependent and independent variables, the instruments, the design, major findings, and conclusions. If pressed for space, concentrate on the problem and, especially, your findings.

Introduction - You will usually start your report with a paragraph or two presenting the investigated problem, the importance of the study, and an overview of your research strategy. You do not need to label this section. Its position within the paper makes that obvious.

The introductory paragraphs are usually followed by a review of the literature. Show how your research builds on prior knowledge by presenting and evaluating what is already known about your research problem. Assume that the readers possess a broad knowledge of the field, but not the cited articles, books and papers. Discuss the findings of works that are pertinent to your specific issue. You usually will not need to elaborate on methods.

The goal of the introduction and literature review is to demonstrate "the logical continuity between previous and present work" (APA, 1994, p. 11). This does not mean you need to provide an exhaustive historical review. Analyze the relationships among the related studies instead of presenting a series of seemingly unrelated abstracts or annotations. The introduction should motivate the study. The reader should understand why the problem was researched and why the study represents a contribution to existing knowledge. Unless the study is an evaluation of a program, it is generally inappropriate to attempt to motivate the study based on its social importance.

Method - The method section includes separate descriptions of the sample, the materials, and the procedures. These are subtitled and may be augmented by further sections, if needed.

Describe your sample with sufficient detail so that it is clear what population(s) the sample represents. A discussion of how the sample was formed is needed for replicability and understanding your study. The APA Task Force on Statistical Inference points out "how a population is defined affects almost every conclusion about an article" (Wilkinson, et al., 1999). Convenience samples are not unusual in scientific inquiry; their use should not discourage you from seeking a publication outlet for your report.

A description of your instruments, including all surveys, tests, questionnaires, interview forms, and other tools used to provide data, should appear in the materials subsection. Evidence of reliability and validity should be presented. Since reliability is a property of scores from a specific use of a specific instrument for a specific population, you should provide reliability estimates based on your data.

The design of the study, whether it is a case study, a survey, a controlled experiment, a meta-analysis, or some other type of research, is conveyed through the procedures subsection. It is here that the activities of the researcher are described, such as what was said to the participants, how groups were formed, what control mechanisms were employed, etc. The description is sufficient if enough detail is present for the reader to replicate the essential elements of the study. It is important for the procedures to conform to ethical criteria for researchers (APA, 1992).

Results - Present a summary of what you found in the results section. Here you should describe the techniques that you used, each analysis and the results of each analysis.

Start with a description of any complications, such as protocol violations and missing data that may have occurred. Examine your data for anomalies, such as outliers, points of high influence, miscoded data, and illogical responses. Use your common sense to evaluate the quality of your data and make adjustments if need be. Describe the process that you used in order to assure your readers that your editing was appropriate and purified rather than skewed your results.

With today's availability of statistical packages, it is fairly easy to use very sophisticated techniques to analyze your data. Understand the techniques you are using and the statistics that you are reporting. Try to use the simplest, appropriate technique for which you can meet the underlying assumptions.

If you are going to use inferential statistics, you should determine the power a priori based on your anticipated distribution, design, and definition of practical significance. This information must stem from your related literature and not the data that you collected. If you fail to reach statistical significance, then this analysis can be used to show that the finding does not stem from low power.

Where appropriate, compute and report effect sizes or, at a minimum, be sure you provide enough information so effect sizes can be computed. Effect sizes provide a common metric for evaluating results across studies and aid in the design of future studies. They will be needed by anyone who attempts a quantitative synthesis of your study along with the others in your area of research.

For most research reports, the results should provide the summary details about what you found rather than an exhaustive listing of every possible analysis and every data point. Use carefully planned tables and graphs. While tables and graphs should be self-explanatory, do not include a table or graph unless it is discussed in the report. Limit them to those that help the reader understand your data as they relate to the investigated problem.

Discussion - At this point, you are the expert on your data set and an authority on the problem you addressed. In this section, discuss and interpret your data for the reader, tell the reader of the implications of your findings and make recommendations. Do not be afraid to state your opinions.

Many authors chose to begin the discussion section by highlighting key results. Return to the specific problem you

investigated and tell the reader what you now think and why. Relate your findings to those of previous studies, by explaining relationships and supporting or disagreeing with what others have found. Describe your logic and draw your conclusions. Be careful, however, not to over generalize your results. Your conclusions should be warranted by your study and your data.

Be sure to recognize the limitations of your study. Try to anticipate the questions a reader will have and suggest what problems should be researched next in order to extend your findings into new areas.

References - There should be a one-to-one match between the references cited in the report and the references listed in the reference section.

PUBLISHING YOUR REPORT

In the process of reviewing the literature, you will have learned which journals publish articles on your topic. If you intend to publish in a journal, these journals will be the most likely candidates. Review the target audience and publication guidelines for these journals to decide which is best suited to your research. Regardless of scholarly quality, a key question in any editor's mind will be whether your manuscript is suited to the journal's purpose and audience. When considering where to submit, note the style of the articles in the journal. For example, if the journal typically publishes articles developing theories based on extensive reviews of the literature and your article is more empirical, then perhaps you should look elsewhere.

Remember that the review process is conducted by human staff, and so is a fallible process. Peters and Ceci (1982) made this point abundantly clear. They retyped just-published articles from prominent journals, and resubmitted them. All of these articles were rejected without it being noticed that they had just been published by the same journals.

Because of high rejection rates and the usual long length of time journals need to make a selection decision, it is tempting to submit a manuscript simultaneously to more than one journal. This, however, is clearly unethical. Most journals appropriately specify that manuscripts under consideration cannot be submitted elsewhere. The editors and reviewers will be taking a considerable amount of time examining your manuscript, usually as volunteers.

You should expect your manuscript to be rejected when it is submitted for the first time. If a manuscript is rejected, you should evaluate the comments and then decide whether to revise, resubmit, or submit it elsewhere. In order to facilitate both your revision and its subsequent evaluation, a resubmission should be accompanied by a description of the issues raised in the review process and your manuscript modifications and other substantive reactions to them.

While very little has been written about ethical standards for authors in the education field, the *Uniform Requirements for Manuscripts Submitted to Biomedical Journals*, which have been adopted by more than 500 scientific and biomedical journals, address critcria for authorship, acknowledgments, redundant publication, competing manuscripts, and conflict of interest. A concise summary of the *Uniform Requirements* can be found in Syrett and Rudner (1996).

A key concept in the *Uniform Requirements* is that individuals identified as authors should have made significant contributions to the conception and design, or analysis and interpretation of data, or both; to drafting of the manuscript or revising it critically for intellectual content; and on final approval of the version of the manuscript to be considered for publication. Being an advisor or head of a research group, does not, in itself, warrant authorship credit.

ADDITIONAL READING

American Psychological Association (1992). Ethical principles of psychologists and code of conduct. *American Psychologist, 47*, 1597-1611. [Available online http://www.apa.org/ethics/code.html].

American Psychological Association (1994). *Publication Manual of the American Psychological Association* (4th ed.). Washington, DC: Author. [amazon]

Gay, L.R. & P.W. Airasian (1999). *Educational Research: Competencies for Analysis and Application.* 6th edition. New York: Prentice Hall.

Peters, D.C., & Ceci, S.J. (1982). Peer review practices of psychological journals: The fate of published articles, submitted again. *The Behavioral and Brain Sciences*, 5, 187-255.

Syrett, Kristen L. & Rudner, Lawrence M. (1996). Authorship Ethics. Practical Assessment, Research & Evaluation, 5(1). [Available online: http://ericae.net/pare/getvn.asp?v=5&n=1].

Thompson, B. (1995). Publishing your research results: Some suggestions and counsel. *Journal of Counseling and Development*, 73, 342-345. [abstract]

Wilkinson, L. and Task Force on Statistical Inference (1999). Statistical Methods in Psychology Journals: Guidelines and Explanations. *American Psychologist*, 54 (8), 594B604. [Available online: http://www.apa.org/journals/amp/amp548594.html].

Descriptors: *Scholarly Journals; *writing for publication;

Home | Articles | Subscribe | Review | Policies

Volume: 8 7 6 5 4 3 2 1

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal. ISSN 1531-7714

Vecchioli, Lisa (1999). A process for evaluating student records management software. *Practical Assessment, Research & Evaluation*, 6(14). Available online: http://ericae.net/pare/getvn.asp?v=6&n=14. This paper has been viewed 9121 times since 12/13/99.

A Process for Evaluating Student Records Management Software

Lisa Vecchioli
Rutgers University

Organizing and managing student records into a cohesive and efficient system might seem like an impossible task. There is a wide array of existing information and information needs, yet schools are often limited by personnel and financial concerns. Large districts can be overwhelmed by the sheer number of students. Further, each institution has its own unique way of keeping track of and reporting on the details of their students' academic and behavioral data. To help schools meet this challenge, several vendors market flexible, high-end software commercial software packages.

Schools need to weigh features and requirements of the software against their own unique needs, desires and capabilities. A condensed version of the introduction to Vecchioli (1998), this article provides an overview of some practical considerations in evaluating such high-end record-keeping software products. Emphasis here is on the evaluation process and the identification of value-based evaluative criteria. A good discussion on some factual criteria for evaluating record-keeping software can be found in Wright (1990).

Evaluations often start with a process to identify the decisions that will be made. An evaluation of record-keeping software should start with a process to identify the individual needs the software product must meet in order to be considered for purchase. How in-depth this process should be depends on the size of the school and the number of officials involved in the decision making process. Large school districts may need to draft detailed requirements and solicit proposal requests from vendors. Small public or private independent schools may need to go through a less formal process. Regardless of size and bureaucratic structure, each school must consider the formation of a school-wide committee, the role the administrator will take in this process, the requirements of the school, the design of the

system, the implementation of the chosen product, and the ability to consult the software company.

Forming a School-Wide Committee & Information Gathering

The first step toward establishing an administrative computing system is to form a school-wide committee that can provide input for developing school-specific evaluation criteria, solicit products from vendors, and examine those products. Connors and Valesky (1986) observe that the primary role of this committee is to identify which school administrative functions are best suited for computerization. This group should consist of a representative group of administrators, teachers, counselors, librarians, and computer experts. Each member should provide input based on their area of expertise.

Administrators should consider involving other future-users who will have the most daily contact with the system (i.e., secretaries, clerks, and business officials). Input from counselors, teachers, and office staff who actually are responsible for scheduling, student record management, creating report cards, and other functions should facilitate the most appropriate software selection. The involvement of these faculty and staff members will familiarize them with the system's structure and capabilities. In turn, these people will be able to take on leadership roles in the computerization of school records by performing such duties as demonstrating particular functions of the software or training other faculty and staff members. Thus, the inclusion of a wide range of people on this committee aids in ensuring the eventual smooth integration of the software into the daily activities of the school.

It is important that all committee members participate in all the evaluation activities. Attending software demonstrations by vendor sales representatives provides a forum for committee members to ask questions regarding their specific areas of expertise. Committee members should also have a chance to use the system or specific module with which they will eventually work. Many vendors provide product demonstrations on CD ROM or on a diskette that users can install on their computer. These product demos enable users to get a sense of what the software interface looks like, how different modules relate to each other, and how specific functions work. If the vendor provides the software on a trial basis, the school may want to consider installing those modules and loading some school data in order to get a better sense of how the system will function in their school setting. Because this is time-consuming for the computer coordinator, schools may want to do this after they have narrowed the decision down to two or three products.

In addition to needs analysis and product evaluation, committee members should be given administrative leave to observe how software packages function at other schools. Regardless of how impressive the sales representative's demonstration is, a demonstration will not be as revealing as seeing how the system functions in an actual school setting. Interviewing other schools can provide committee members with a greater understanding of how the system can increase their own school's productivity as well as what initial training and data entry tasks they face.

Needs Analysis

Once a committee has been established, the members should examine which administrative functions need to be computerized. The software packages generally consist of modules that can be purchased separately and address particular functions such as school records, attendance, scheduling, and progress (grades or marks). The committee might begin by examining the current management process of these areas and deciding what functions could be expedited by automation and how the software must be able to accommodate for the school's particular method of representing data. For example, the software system must be able to adapt to how the school calculates grades as well as how the school creates and formats its schedule. Most reputable vendors provide enough flexibility in their programs to allow for user-defined fields and a variety of scenarios. However, if a specific need cannot be met by the software packages under evaluation, schools may have to make some concessions. If a school's unique needs are

known before the software is purchased, accommodations usually can be made.

As schools begin to delineate how data is currently gathered and used in the four broad areas addressed by integrated student records management software (school records, grades, scheduling, and attendance), they should also begin to establish a priority order for the integration of the selected software product. It may not be advisable to automate all areas. For example, the needs of a small private elementary school may not warrant the purchase of the most powerful scheduling module available in integrated systems. It may be easier for a school like that to use a generic database program and import data into that program to create a student's schedule.

After the committee has determined which data management areas need to be computerized, the committee should prioritize in what order these areas should be addressed. The school should be guided by three main factors during this process: the availability of finances, the needs of the school, and the ability to train school personnel. School funding determines which software systems and which modules a school committee should consider during the decision process. The ability to train people to use the system determines how effectively the system will be used. This is especially troublesome if the software is complex. Therefore, a committee should consider how well its school can prepare all faculty and staff members who need to access and input data in the system to perform required tasks.

The two fundamental issues that a school committee should examine are the school's capability to enable faculty and staff to make more productive use of their time and its capability to provide accurate data on individual students that can be utilized in a way which effectively meets the needs of the students. The data gathering and reporting abilities of integrated systems allow school personnel to create fuller descriptions of individual students' progress and achievement than was previously possible using traditional reports. Additionally, student records software provides a greater variety of comment, increases pupil involvement in and responsibility for the reporting process, assists the integration of curriculum and good pedagogic practice, and produces a more constructive and positive diagnostic assessment of pupil progress (Wilson and Armstrong, 1993).

The Role of the Administrator

Integrated student records management systems allow for more efficient organization of school data. Powerful reporting and query capabilities permit administrators to track and analyze data in ways that were not previously possible. Moreover, integrated software packages give school or building-level administrators within districts more independence for gathering and analyzing data. These also keep the administrators from being "completely dependent on the services of a central or district data processing manager" (Bozeman and Spuck, 1994, p.42).

Knowing the administrative importance of choosing an effective integrated student records management software product, it is clear to see that school officials need to play a vital role in deciding which administrative function should be automated. Administrators are able to provide important information about their school's current and future record keeping needs. Moreover, school administrators determine the degree to which a software product will be utilized in order to "contribute to institutional improvement" (Bers, 1992, p. 3).

System Design

As the working committee members gather information about packages, they should consider how the system will ultimately serve their unique and general institutional needs. A set of criteria should be drafted in order to compare and evaluate each system. Peter Wright (1990) suggested "staged approach" to evaluation where "systems are evaluated against progressively more detailed criteria" (p.218). The first stage of evaluation is characterized by the performance of certain tasks:

- the identification of software products,
- the acquisition of information such as literature reviews,
- discussions with product developers / vendors as well as the faculty and staff of other schools who use different software products,
- the general screening of available software, and
- the analysis of institutional needs.

During this stage of evaluation, committee members should partake in system demonstrations and detailed discussions with developers/distributers. Wright (1990) advised that as the evaluators think about how the system will meet their particular needs, the resulting analysis should be a reflection of the following:

- Current needs and requirements (i.e. the manner in which things are presently done)
- How things should operate in the future
- Potential uses of the system that committee members previously did not know were possible

Throughout this process, analysis will shift from general system considerations to the module specific criteria. The knowledge and expertise of individual committee members will be invaluable as the analysis begins to narrow in focus.

Once this stage is complete, the committee members should be able to recommend a system that will meet most if not all of their school's current and anticipated needs. The decision should be based on sufficient data and information as well as a thorough analysis of available software products. However, if a final decision is not imminent at this point, the committee members might consider developing and using a quantitative measure on which they can base their decision. This process includes assigning weighted scores to both the general systems and module specific criteria as well as calculating the performance of each system based on a ration of how well a system performed compared to how well it could have performed. While this approach is certainly more objective than using a "checklist" procedure, it is probably too time consuming for the members of evaluation committees who are involved in this process in addition to teaching and administrative responsibilities. As Wright (1990) indicates, this process is more suited for districts or consortia of private independent schools that have time and resources.

References

Bers, T. H.. (1992). Of student tracking systems. *Community, Technical and Junior College Journal. 62*(4), 21-23.

Bozeman, W. C. and Spuck, D. W. (1994). "Computer support for administrative leadership of schools." In Greg Kearsley & William Lynch (Eds.) *Educational Technology: Leadership Perspectives.* (pp. 39-50). Englewood Cliffs, NJ: Educational Technology Publications.

Connors, E. T. and Valesky, T. C. (1986). *Using Microcomputers in School Administration. Fastback no. 248.* Bloomington, IN: Phi Delta Kappa Educational Foundation.

Vecchioli, L. (1998). *Evaluating Student Records Management Software.* College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Wilson, B. and D. Armstrong (1993). A Computerized System for School Report and Record Writing. *Computers and Education: An International Journal, 21*(4), 321-330.

Wright, P. (1990). Choosing a computer based instructional support system: An evaluation / selection model" *Computers and Education, 14*(3), 217-225.

Descriptors: Computer Software Evaluation; *Computer Software Selection; Computer Uses in Education; Criteria; Educational Administration; *Evaluation Methods; *Records Management; *Student Records

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").